Essays on the Macroeconomic Effects of Imperfect Banking Competition and Other Financial Frictions

Jiaqi Li

Faculty of Economics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Queens' College

July 2019

I would like to dedicate this thesis to my loving parents, Yue Yang and Zhonglu Li.

Declaration

I hereby declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit of 60,000 words.

Jiaqi Li
July 2019

Acknowledgements

I am deeply indebted to my supervisor, Petra Geraats, for her continued guidance and invaluable advice. Throughout my time as a Ph.D. student, I have benefited greatly from the discussions with Petra. Her intellectual drive and rigorous attitude towards research have inspired me immensely. There is no doubt that the things I have learned from her will benefit me for a lifetime and I will forever be grateful for her support and help throughout these years. I am also indebted to my advisor, Donald Robertson, for his valuable feedback on my empirical work and his continued support.

I am extremely grateful to my supervisor at the International Monetary Fund, Jérôme Vandenbussche, who provided me with the Orbis firm-level data to carry out the empirical analysis in my last chapter and also gave me some insightful feedback. The final chapter of this thesis was inspired from the project on the riskiness of credit allocation during my internship at the IMF. This chapter would not exist without his help.

My research has benefited greatly from conversations with many others, including: Adrien Auclert, Charles Brendon, Vasco Carvalho, Tiago Cavalcanti, Giancarlo Corsetti, Elisa Faraglia, Chryssi Giannitsarou, Pete Klenow and Weilong Zhang. I am also grateful to seminar and conference participants at a number of locations, including the University of Cambridge, the Bank of Canada, International Finance and Banking Society, and Money, Macro and Finance Ph.D. Workshop.

I am grateful for the support and help that I have received from my friends and Ph.D. colleagues at Cambridge. Especially, I want to thank my boyfriend, Lu Han, who has given me many helpful suggestions for my research and accompanied me through the most difficult times during my Ph.D. I am also thankful of the help that I have received from River Chen, Sihua Ding, Simon Lloyd, Margit Reischer and Jasmine Xiao, as well as Anil Ari, Shuyi Ge, Zeina Hasna, Weiguang Liu, Monica Petrescu, Dan Wales and Peng Zhang.

I am thankful to my Economics teacher at Oxford High School, Mike Gallacher, and my undergraduate tutors at Oxford, Brian A'Hearn and George Bitsakakis, for introducing me to Economics and encouraging me to continue my studies to this level.

Most of all, I am indebted to my parents, Yue Yang and Zhonglu Li, who have always supported me in pursuing my dream, both spiritually and financially.

Abstract

In this thesis, I study the effects of financial frictions and in particular, imperfect banking competition, on different macroeconomic aspects. The thesis consists of a short introductory chapter and three papers.

The first paper investigates the impact of imperfect banking competition on aggregate fluctuations in a dynamic stochastic general equilibrium (DSGE) framework. Following the global financial crisis, there has been an increasing focus on incorporating financial frictions into a DSGE model, often by introducing an agency problem which serves to amplify macroeconomic shocks. This paper examines the impact of another important financial friction, imperfect competition in banking, on aggregate fluctuations by incorporating a Cournot banking sector into a DSGE model that features an agency problem that gives rise to collateral constraints. In the presence of a binding collateral constraint, imperfect banking competition is found to have an amplification effect on aggregate fluctuations after a contractionary monetary policy shock and adverse collateral shocks. Adverse shocks that make borrowers more financially constrained and their loan demand more inelastic can induce banks with market power to raise the loan rate, resulting in a countercyclical loan interest margin that amplifies aggregate fluctuations.

The second paper studies how imperfect competition in the banking sector affects financial stability. By building a model of imperfect banking competition featuring the accumulation of bank equity via retained earnings, I find that bank competition can have different short-run and long-run effects on financial stability. In the short run, less competition can jeopardize stability as it increases banks' loan assets and thus lowers their equity-to-assets ratios (equity ratios), making them more likely to default. In the long run, less competition tends to enhance stability as banks make higher profits and accumulate equity faster over time, resulting in higher equity ratios and hence lower bank default probabilities. The extent of this long-run stability gain from less competition and whether the stability gain outweighs the efficiency loss crucially depend on banks' dividend distribution or macroprudential policies. Empirically, I find two sets of supporting evidence for the model predictions using a large bank-level panel from EU and OECD countries spanning around 15 years. First, bank concentration, an inverse measure for competition, has a significant positive effect on the change in bank equity. Second, banks' equity ratios are found to be negatively related to their default probabilities, which are proxied by credit default swap spreads.

In the third paper, I study the impact of financial frictions in the form of borrowing constraints on the efficient allocation of physical capital. While it is widely perceived that financial frictions have adverse impact on capital allocation, the importance of this impact is difficult to quantify. This paper presents a novel two-step approach to estimate the importance of financial frictions on capital misallocation, measured by the dispersion of the marginal revenue product of capital. First, based on the general theoretical result that the capital investment of financially constrained firms is more sensitive to their internal financing than for unconstrained firms, I use a switching regression approach to jointly estimate the two different investment regimes and the probability of each firm being constrained. Firms are classified as financially constrained or unconstrained based on the estimated probabilities. Second, I provide a decomposition of capital misallocation and estimate the fraction that can be explained by the presence of financially constrained firms. Applying this method to large panels of manufacturing firms for 20 countries from the 1990s to 2015, this paper finds that for most countries and two-digit industries, more than a quarter of firms are classified as financially constrained. Furthermore, the presence of these constrained firms accounts for more than half of capital misallocation.

Contents

Chapter 1

Introduction

The global financial crisis has highlighted the impact of financial intermediaries on the real economy. This suggests that the financial sector is not frictionless and it is important to incorporate financial frictions in the study of aggregate fluctuations, because they can cause shocks and affect the propagation of other macroeconomic shocks. So far, the most commonly used financial frictions in the literature are agency problems between lenders and borrowers (e.g., Gertler et al., 2012; Gertler and Karadi, 2011; Gertler et al., 2010; Gilchrist et al., 2009; Iacoviello, 2005; Bernanke et al., 1999; Carlstrom and Fuerst, 1997; Kiyotaki and Moore, 1997). As agents' balance sheet conditions worsen during bad times, indicating more severe agency problems, the resulting increased difficulty in obtaining external finance tends to amplify the initial shock that adversely affects balance sheet conditions (Bernanke et al., 1996).

One special feature of the banking sector is that it tends to be imperfectly competitive and highly concentrated. Despite a large empirical literature documenting the market power of the banking sector (e.g., Bikker and Haaf, 2002; Ehrmann et al., 2001; De Bandt and Davis, 2000; Oxenstierna, 1999; Berg and Kim, 1998), this financial friction is often neglected in the study of macroeconomic fluctuations. Chapter 2 investigates whether imperfect banking competition affects aggregate fluctuations. By incorporating both imperfect banking competition and an agency problem into a dynamic stochastic general equilibrium (DSGE) framework, Chapter 2 addresses how imperfect banking competition interacts with the agency problem and whether it can also amplify aggregate fluctuations on top of the agency problem.

After the financial crisis, central banks around the world have taken actions to safeguard financial stability to prevent future crises, such as implementing various macroprudential policies. A natural question following from Chapter 2 is whether imperfect banking competition also affects financial stability. The answer to this question can provide crucial guidance on choosing the most effective macroprudential policy tools. Despite its importance, the relationship between banking competition and financial stability remains highly debated in the literature (e.g., Corbae and Levine, 2018; Martinez-Miera and Repullo, 2010; Boyd and De Nicolo, 2005; Allen and Gale, 2000; Keeley, 1990).

Chapter 3 investigates to what extent imperfect banking competition could enhance financial stability by making banks less likely to default due to higher profits and how it could lead to a trade-off between financial stability and macroeconomic efficiency.

Shifting the focus from macroeconomic efficiency to allocative efficiency, Chapter 4 studies the efficient allocation of physical capital across producers in the presence of financial frictions. Since capital misallocation has adverse implications on aggregate productivity, understanding the causes of it has become one of the central topics in the recent literature (e.g., Bai et al., 2018; David and Venkateswaran, 2017; Gopinath et al., 2017; Midrigan and Xu, 2014; Bartelsman et al., 2013; Gilchrist et al., 2013; Restuccia and Rogerson, 2013). Financial frictions are often regarded as a contributing factor for capital misallocation, but the magnitude of their impact is difficult to quantify. Chapter 4 focuses on financial frictions in the form of collateral constraints and provides a new method to quantify the impact of financial frictions on capital misallocation by using large firm-level datasets across countries. I then apply this method to estimate the fraction of capital misallocation that can be explained by the presence of financially constrained firms.

This thesis examines the macroeconomic effects of imperfect banking competition and other financial frictions. The next two chapters study the effects of imperfect banking competition on macroeconomic volatility and financial stability, respectively. I then focus on the effect of other financial frictions on the efficient allocation of physical capital in the final chapter.

In Chapter 2, I study how imperfect banking competition affects macroeconomic fluctuations by incorporating a Cournot banking sector into a DSGE framework embedded with an agency problem that gives rise to collateral constraints. I use the Cournot model to characterise oligopolistic banking competition since the banking sector is highly concentrated and is often dominated by a few large players. For instance, in most EU and OECD countries, the largest five banks account for more than 60% of the market share (using ECB and Bankscope data in 2007 and 2014).

I find that in the presence of a binding collateral constraint, imperfect banking competition can amplify aggregate fluctuations via a countercyclical loan interest margin, which refers to the difference between the loan rate and the deposit rate. For instance, after a contractionary monetary policy shock, the binding constraint tightens and borrowers are more financially constrained. As a result, loan demand becomes more inelastic, which induces banks with market power to charge a higher loan rate, leading to a countercyclical loan margin that is large enough to amplify aggregate fluctuations. The countercyclical loan interest margin is documented in Chapter 2 using country-level data for EU countries from 1980 to 2016.

Some papers have incorporated imperfect banking competition into a DSGE model, typically using monopolistic competition within the Dixit and Stiglitz (1977) framework (Cuciniello and Signoretti, 2015; Hafstead and Smith, 2012; Dib, 2010; Gerali et al., 2010;

Hülsewig et al., 2009). However, assuming agents demand a composite bundle of loan and deposit contracts from many different banks is unrealistic, given that in reality, households and firms tend to rely on only one bank or at most a few. Chapter 2 avoids this assumption by using a Cournot banking sector. More importantly, the banking sector tends to be very concentrated and dominated by a few large banks, which implies that oligopolistic competition such as Cournot may be more appropriate. Another contribution of Chapter 2 is that I find imperfect banking competition can amplify the response of output via a countercyclical loan interest margin after a monetary policy shock, which differs from the attenuation effect often found in the existing literature (Andrés and Arce, 2012; Hafstead and Smith, 2012; Dib, 2010; Gerali et al., 2010). Although Cuciniello and Signoretti (2015) also find that imperfect banking competition can amplify aggregate fluctuations after a monetary policy shock, their results rely on the strategic interaction between banks with market power and the inflation-targeting central bank, which cannot be applied to countries that have a fixed exchange rate regime or are part of a large monetary union (such as the eurozone).

Chapter 3 studies how imperfect banking competition affects financial stability. Much of the literature has focused on how banking competition affects banks' or borrowers' risk-taking (e.g., Corbae and Levine, 2018; Boyd and De Nicolo, 2005; Allen and Gale, 2000; Keeley, 1990). Instead, Chapter 3 studies how competition affects banks' equity-to-assets ratios (equity ratios) and thereby financial stability measured through banks' default probabilities. By building a model of Cournot banking competition featuring the accumulation of bank equity via retained earnings, I find that less banking competition can lead to a large gain in financial stability provided that banks retain the greater profits to build up their capital buffer.

Although imperfect banking competition can enhance financial stability, it leads to a higher loan rate and thereby a lower demand for physical capital, which reduces aggregate output and hence macroeconomic efficiency. Another contribution of Chapter 3 is to quantify the importance of the financial stability gain from imperfect banking competition relative to the macroeconomic efficiency loss. In doing so, I find that bank equity accumulation is important for understanding the trade-off between financial stability and macroeconomic efficiency. In the absence of bank equity accumulation, the financial stability gain from imperfect banking competition is very limited and is outweighed by the macroeconomic efficiency loss. As a result, perfect banking competition is the best in this case. However, when banks retain their profits to build up their capital buffer over time, for instance, in response to macroprudential regulations, the financial stability gain can become large enough to outweigh the macroeconomic efficiency loss. Chapter 3 also provides new empirical evidence by assessing the model prediction that in the presence of bank equity accumulation, less banking competition improves financial stability, using a large bank-level panel of EU and OECD countries over the period from 1999 to 2016.

In Chapter 4, I study the effect of other financial frictions in the form of collateral constraints on capital misallocation measured by the dispersion of the marginal revenue product of capital (MRPK). I provide a novel two-step approach to estimate the fraction of the dispersion of MRPK that is caused by the presence of financially constrained firms, which relies on few restrictive assumptions and can be readily applied to a large number of countries. First, based on the theoretical result that the capital investment of financially constrained firms is more sensitive to their internal financing than for unconstrained firms, I use a switching regression approach to jointly estimate the two different investment regimes and the probability of each firm being constrained. Firms are classified as financially constrained or unconstrained based on the estimated probabilities. Second, I decompose the dispersion of MRPK across all firms into the dispersions and means of MRPK for the two types of firms. Using the firm classification and the decomposition, the fraction of the dispersion of MRPK caused by the presence of financially constrained firms can then be estimated.

Applying this method to large panels of manufacturing firms for 20 countries from the 1990s to 2015, I find that the dispersions and means of MRPK for the financially constrained types are much larger than those for the unconstrained firms. Furthermore, for most countries and two-digit industries, more than a quarter of firms are classified as financially constrained and the presence of these constrained firms accounts for more than half of capital misallocation.

To conclude, the thesis has shown the importance of imperfect banking competition for macroeconomic fluctuations and financial stability, and has quantified the impact of financially constrained firms on the allocation of physical capital.

Chapter 2

Imperfect Banking Competition and Macroeconomic Volatility: A DSGE Framework

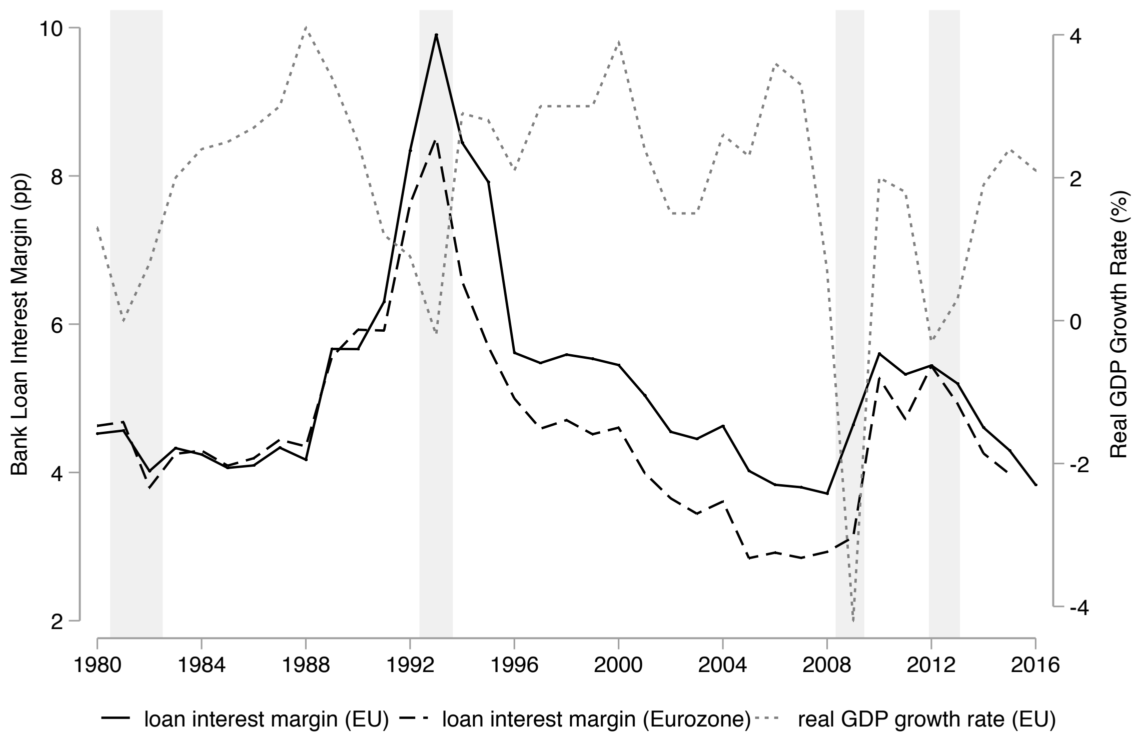
2.1 Introduction

Following the global financial crisis, there has been an increasing focus on incorporating financial frictions into a dynamic stochastic general equilibrium (DSGE) model. Most of the existing literature studying the role of financial frictions in amplifying aggregate fluctuations often models the financial friction using an agency problem and assumes a perfectly competitive banking sector. However, the banking sector tends to be imperfectly competitive in reality. For instance, in most EU and OECD countries, the largest five banks account for more than 60% of the market share (using ECB and Bankscope data in 2007 and 2014), suggesting that the banking sector tends to be dominated by a few large players.[1] Furthermore, neglecting imperfect banking competition may miss out an important propagation mechanism of macroeconomic shocks.

This paper incorporates a Cournot banking sector into a DSGE model embedded with an agency problem that gives rise to collateral constraints. I find that in the presence of a binding collateral constraint, imperfect banking competition can lead to a countercyclical loan interest margin (which rises during bad times) that amplifies aggregate fluctuations after monetary policy shocks and collateral shocks. In this paper, the loan interest margin refers to the difference between the loan rate and the deposit rate. With perfect banking competition and no other frictions, this loan interest margin equals zero and thus the loan rate moves one-to-one with the deposit rate. By contrast, with imperfect banking competition, the loan interest margin is endogenously changing, which is consistent with

[1]See Oxenstierna (1999) and Berg and Kim (1998) for empirical evidence of oligopolistic banking competition in Sweden and Norway respectively.

Figure 2.1: Bank Loan Interest Margin and Real GDP Growth in EU from 1980 to 2016



Note: The annual bank loan interest margin (in percent points) from the World Bank is the difference between the lending rate (charged by banks on loans to the private sector) and deposit rate (offered by commercial banks on three-month deposits). The graph plots the unweighted average loan interest margin across EU/Eurozone countries over time. The dotted line corresponds to the annual real GDP growth rate for the EU. The shaded area represents the Euro area recessions (i.e., 1980Q2-1982Q3, 1992Q1-1993Q3, 2008Q2-2009Q2, 2011Q4-2013Q1) documented by the Center for Economic Policy Research (https://cepr.org/content/euro-area-business-cycle-dating-committee).
Data sources: World Bank, IMF World Economic Outlook

empirical evidence. Figure 2.1 documents the countercyclical loan interest margin using country-level data for EU countries from 1980 to 2016.

More specifically, how the loan margin changes in response to shocks crucially depends on the elasticity of loan demand to the loan rate. Based on the model, the loan demand becomes more inelastic when the expected future prices of capital and housing decrease and/or the expected marginal products of capital and housing increase. Intuitively, in the presence of a binding collateral constraint, lower expected asset prices indicate reduced borrowing capacity and higher expected marginal products for the borrowers indicate that they operate further below the optimal scale due to a more tightly binding borrowing constraint.[2] Hence, lower expected asset prices and higher expected marginal products can both imply that the binding borrowing constraint tightens and hence borrowers are more financially constrained, leading to a more inelastic loan demand. Furthermore, after a negative shock to the fraction of the collateral value which the borrower can borrow

[2]The relation between the binding borrowing constraint and marginal products is further analysed in Chapter 4.

against (i.e., a negative collateral shock), the reduction in this fraction can also reduce the loan demand elasticity by making borrowers more financially constrained. The lower elasticity of loan demand gives banks with market power an incentive to charge a higher loan rate, leading to a greater loan interest margin for a given deposit rate.

This paper finds that after a contractionary monetary policy shock and negative collateral shocks, borrowers are more financially constrained, implied by a lower leverage ratio and a tightening of the binding collateral constraint. As a result, loan demand is more inelastic, which leads to a rise in the loan margin that amplifies the loss in output. After a negative productivity shock, however, the amplification effect is much weaker in spite of a countercyclical real loan margin. This is because a negative productivity shock is inflationary, so the real loan rate falls, which tends to reduce the tightness of the binding constraint, *ceteris paribus*. In other words, since debt is denominated in nominal terms, a higher price level reduces the real debt burden and this debt-deflation effect tends to dampen the aggregate fluctuations.

This paper contributes to the literature on DSGE modelling with financial frictions. This literature can be broadly divided into three different strands: agency problems, time-varying loan spreads, and imperfect banking competition.

The first strand of the literature only incorporates an agency problem between borrowers and lenders into a DSGE model to generate the financial accelerator effect (Bernanke et al., 1996). As borrowers' balance sheet conditions worsen during bad times, agency problems become more severe, and the resulting increased difficulty in obtaining external finance tends to amplify any shocks that adversely affect balance sheet conditions. The agency problem is often modelled by costly debt enforcement (Gertler et al., 2012; Gertler and Karadi, 2011; Gertler et al., 2010; Iacoviello, 2005; Kiyotaki and Moore, 1997). As borrowers cannot be forced to repay unsecured debt (Beck et al., 2014), creditors would not lend an amount that exceeds the value of collateralized assets and hence borrowers face a collateral constraint. Alternatively, the agency problem can be modelled by the costly state verification of Townsend (1979) that leads to an endogenous external finance premium (EFP), which then raises the cost of borrowing and amplifies business cycle fluctuations (Agénor and Montiel, 2015; Gilchrist et al., 2009; Bernanke et al., 1999; Carlstrom and Fuerst, 1997).

In all these papers where the financial intermediation is explicitly modelled, the banking sector is assumed to be perfectly competitive. However, there is a large empirical literature on banking competition and a common finding is that banks indeed have market power, and that competition levels vary across countries and over time (Bikker and Haaf, 2002; Ehrmann et al., 2001; De Bandt and Davis, 2000). This paper incorporates an imperfectly competitive banking sector into a DSGE model with collateral constraints.

The second strand of the literature focuses on the role of a time-varying loan spread in understanding business cycle fluctuations (Cúrdia and Woodford, 2015; Gertler and Karadi, 2011; Gertler et al., 2010; Gilchrist et al., 2009; Goodfriend and McCallum, 2007; Bernanke and Gertler, 1989), using a variety of different assumptions (other than imperfect banking competition) to generate the spread. For instance, Cúrdia and Woodford (2015) introduce a time-varying spread by assuming that the loan-origination process would consume real resources and that there is an exogenously varying loss rate on loans. Similarly, Goodfriend and McCallum (2007) assume bank loans and deposits are produced by a competitive banking sector according to a Cobb-Douglas production function. The costly production process gives rise to an EFP, which can be procyclical or countercyclical in response to a monetary shock depending on different parameter calibrations. Instead, in this paper, I use imperfect banking competition to generate the time-varying loan interest margin.

My paper is closely related to the third strand of the literature that focuses on incorporating imperfect banking competition into a DSGE model. In the existing literature, imperfect banking competition is often modelled via monopolistic competition within the Dixit and Stiglitz (1977) framework (Hafstead and Smith, 2012; Dib, 2010; Gerali et al., 2010; Hülsewig et al., 2009). However, assuming that agents demand a composite basket of loan and deposit contracts with constant elasticity of substitution (CES) is unrealistic, given that in practice, firms and households tend to rely largely on one bank or at most a few different banks. Furthermore, Dixit and Stiglitz (1977)'s model of monopolistic competition results in a constant loan margin, which implies that the loan rate moves one-to-one with the deposit rate set by the central bank.[3] However, the loan margin is not a constant and it changes over the business cycle, as shown in Figure 2.1. To generate an endogenously changing loan margin from imperfect banking competition, Andrés and Arce (2012) use Salop's (1979) model of monopolistic competition and Cuciniello and Signoretti (2015) introduce large banks into the Dixit and Stiglitz (1977) framework. Both papers find a countercyclical loan margin, but in the former case, it is not large enough to amplify aggregate fluctuations after a contractionary monetary policy shock. Although Cuciniello and Signoretti (2015) find that monopolistic banking competition can amplify aggregate fluctuations after a contractionary monetary policy shock, their result relies on the strategic interaction between banks with market power and an inflation-targeting central bank, which cannot explain the empirically observed countercyclical loan interest margin in countries that have a fixed exchange rate regime or are part of a large monetary union (such as the eurozone).[4]

[3]In all these papers, changes in the loan margin over the business cycle are generated by introducing exogenous shocks to the elasticity of substitution between different loan or deposit products (e.g., Gerali et al., 2010) or bank's marginal cost of producing loans (e.g., Hafstead and Smith, 2012), or modelling interest rate stickiness *à la* Calvo (1983) or Rotemberg (1982) (e.g., Dib, 2010; Gerali et al., 2010; Hülsewig et al., 2009), other than imperfect banking competition.

[4]Using the country-level loan interest margins from the World Bank and the real GDP growth rates from the IMF for EU countries from 1980 to 2016, the countercyclical loan interest margin also holds in small eurozone countries, such as Belgium, Estonia, Finland, and Netherlands. As a result, they do not have independent monetary policy and since they are small, it is unlikely for them to influence the monetary policy set by the ECB.

This paper contributes to this third strand of the literature in two ways. First, imperfect banking competition is modelled by a Cournot banking sector, which avoids the unrealistic assumption on agents' preferences for banking and the unrealistic fact that the loan margin remains constant over the business cycle in the framework of monopolistic competition *à la* Dixit and Stiglitz (1977). Furthermore, since banking sector tends to be very concentrated and dominated by a few large banks, Cournot model captures the oligopolistic competition, and at the same time, also nests perfect banking competition as a special case (which is used as a benchmark in this paper). To the best of my knowledge, this is the first paper that incorporates Cournot banking competition into a DSGE framework.

Second, in this paper, the binding borrowing constraint is tied to the values of both housing and physical capital collateral, while the existing studies in this strand of the literature only model one type of collateral asset, using either housing or capital (e.g., Cuciniello and Signoretti, 2015; Andrés and Arce, 2012; Gerali et al., 2010). In doing so, I find that imperfect banking competition can amplify aggregate fluctuations via a countercyclical loan interest margin after a contractionary monetary policy shock, which differs from the attenuation effect often found in the existing literature (Andrés and Arce, 2012; Hafstead and Smith, 2012; Dib, 2010; Gerali et al., 2010). This is because housing is assumed to be in inelastic supply, thus any reduction in housing demand leads to a large fall in its price, which in turn reduces borrowers' value of housing collateral and hence their borrowing capacity. As borrowers become more financially constrained, the loan demand becomes more inelastic, inducing banks with market power to charge a higher loan rate. In addition, the presence of capital collateral in the binding borrowing constraint also contributes to the amplification effect since any reduction in its demand and price will also reduce borrowers' borrowing capacity and make them more financially constrained. As a result, the countercyclical loan margin is large enough to amplify aggregate fluctuations after a contractionary monetary policy shock in this paper.

The remainder of the paper is structured as follows. Section 2.2 introduces the model to analyse the effect of imperfect banking competition in a DSGE model embedded with a collateral constraint. Section 2.3 explains the calibration of model parameters. Section 2.4 shows the impulse responses of some key variables after a contractionary monetary shock, a negative productivity shock and collateral shocks. Section 2.5 discusses robustness checks, and Section 2.6 concludes.

2.2 The Model

The model aims to show the effect of imperfect banking competition relative to perfect banking competition on aggregate fluctuations in a framework of collateral constraints. Section 2.2.1 shows the model set-up for perfect banking competition and Section 2.2.2 replaces the perfectly competitive banking sector with a Cournot banking sector.

2.2.1 Perfect Banking Competition Benchmark

There are six types of agents: households, entrepreneurs, retailers, capital producers, banks, and a central bank. Each of the former five agent types has a unit mass. There is a fixed housing supply that can be invested by households and entrepreneurs, following Iacoviello (2005) and Andrés and Arce (2012). Households consume, supply labor to the entrepreneurs, invest in housing and decide how much to save via one-period non-state-contingent nominal bank deposit contracts or one-period risk-free nominal bonds. Perfectly competitive entrepreneurs are born with some physical capital and housing in the initial period and they have access to a Cobb-Douglas production technology. They hire labor from households, purchase new capital from capital producers and purchase real estate from the households to produce a wholesale good. The wholesale good produced by entrepreneurs cannot be consumed directly and is sold to monopolistically competitive retailers who then differentiate the wholesale good costlessly into different varieties. Each retailer uses the wholesale good as the only input to produce a different variety. The final consumption good is a composite CES (constant elasticity of substitution) bundle of all the varieties. Perfectly competitive capital producers buy the undepreciated capital from entrepreneurs and consumption goods from retailers to produce new capital, which is then sold back to the entrepreneurs.

Banks offer two types of one-period contracts: deposit contracts and loan contracts. The contracts are denominated in nominal terms, which means they are not inflation-indexed and the borrowing or saving decisions are made on the basis of a preset contractual nominal loan or deposit rate. Assuming nominal bank deposits and one-period riskless nominal bonds are perfect substitutes to households under full deposit insurance, the gross nominal deposit rate must equal the gross nominal interest rate R_t earned on the riskless nominal bond invested in period t. Following Andrés and Arce (2012), this paper abstracts away from the deposit insurance premium in the banking sector's problem. Since banks are perfectly competitive, each of them takes the nominal loan rate as given and maximizes its profit with respect to the loan (or deposit) quantity. Assuming costless financial intermediation and no expected default on loans,[5] the gross nominal loan rate $R_{b,t}$ equals the gross nominal deposit rate R_t, which is controlled by the central bank.

Households

There is a continuum of identical infinitely-lived households of unit mass. The representative household maximizes the following expected utility:

$$E_t \sum_{s=0}^{\infty} \beta^s [ln(c_{t+s}) + \phi_l ln(1 - l_{t+s}) + \phi_h ln(h_{t+s})] \tag{2.1}$$

[5] Under reasonable calibration, the steady state net worth of the entrepreneur is large enough that even after a very large and persistent negative productivity shock (five times its standard deviation), the net worth is far from being negative and the entrepreneur is able to afford the gross loan interest payment. So this paper ignores the possibility of default on loans.

which depends on consumption c, labor supply l and real estate holdings h, with E_t being the expectation operator conditional on information in period t, and $\beta \in (0, 1)$ the subjective discount factor of the household. The total time endowment is normalised to 1, so $(1 - l_t)$ denotes the amount of period-t leisure time, and $\phi_l > 0$ and $\phi_h > 0$ are the relative utility weights on leisure time and housing respectively. As in Gertler and Karadi (2011), a cashless economy is considered here for the convenience of neglecting real money balances in the utility function.

In each period t, the household consumes c_t, saves d_t in real (final consumption) terms, invests in housing h_t and supplies labor hours l_t. Assume there is zero net supply of risk-free nominal bonds, so in equilibrium, households hold only nominal bank deposits. The nominal deposits d_{t-1} saved in period $t-1$ earn a gross nominal interest rate R_{t-1} at the beginning of period t. Let p_t denote the unit price of the final consumption good, then the gross inflation rate is $\pi_t \equiv \frac{p_t}{p_{t-1}}$. Assume retailers, capital producers and banks are owned by the households. Given the gross real interest earnings on deposits $\frac{R_{t-1}d_{t-1}}{\pi_t}$ at the beginning of period t, real labor income $w_t l_t$ and real lump-sum profits Π_t^R, Π_t^{CP} and Π_t^B made by retailers, capital producers and the banking sector respectively, the household decides how much to consume and save and how much housing investment $(h_t - h_{t-1})$ to make in period t. Assuming there is no depreciation of housing, the representative household faces the following budget constraint:

$$c_t + d_t + q_{h,t}(h_t - h_{t-1}) = \frac{R_{t-1}d_{t-1}}{\pi_t} + w_t l_t + \Pi_t^R + \Pi_t^{CP} + \Pi_t^B \tag{2.2}$$

where $q_{h,t}$ is the real price of housing. Let λ_t denote the Lagrange multiplier associated with the budget constraint or equivalently, the marginal utility of consumption. The first order conditions with respect to consumption c_t (2.3), labor supply l_t (2.4), housing h_t (2.5), and bank deposits d_t (2.6) are as follows:

$$\lambda_t = \frac{1}{c_t} \tag{2.3}$$

$$\frac{\phi_l}{1 - l_t} = \lambda_t w_t \tag{2.4}$$

$$\frac{\phi_h}{h_t} + \beta E_t[\lambda_{t+1} q_{h,t+1}] = \lambda_t q_{h,t} \tag{2.5}$$

$$\lambda_t = \beta E_t \left[\lambda_{t+1} \frac{R_t}{\pi_{t+1}} \right] \tag{2.6}$$

Equation (2.6) is the standard intertemporal Euler equation, which can also be written as:

$$1 = E_t \left[\Lambda_{t,t+1} \frac{R_t}{\pi_{t+1}} \right] \tag{2.7}$$

where $\Lambda_{t,t+1} \equiv \beta\frac{\lambda_{t+1}}{\lambda_t} = \beta\frac{u'(c_{t+1})}{u'(c_t)}$ is the stochastic discount factor in period t for real payoffs in period $t+1$, with $u(c) = ln(c)$.

Entrepreneurs

Assume entrepreneurs are born with some physical capital and housing in the initial period. In period $t-1$, a continuum of perfectly competitive entrepreneurs of unit mass acquire physical capital k^E_{t-1} from capital producers at the real price q_{t-1} and real estate h^E_{t-1} from households at the real price $q_{h,t-1}$ for production in period t. Capital k^E_{t-1}, housing h^E_{t-1} and labor l^E_t hired from households are used as inputs to produce the wholesale good in period t using a constant-returns-to-scale Cobb-Douglas production technology:

$$y_{w,t} = z_t(k^E_{t-1})^{\alpha}(h^E_{t-1})^{v}(l^E_t)^{1-\alpha-v} \tag{2.8}$$

where $\alpha \in (0,1)$ and $v \in (0,1)$ are the output elasticities of physical capital and housing respectively and $0 < \alpha + v < 1$. $y_{w,t}$ is the output of the wholesale good (which differs from the output of the final consumption good by a factor of the price dispersion as will be shown by equation (2.38) in Section 2.2.1). Productivity z_t follows an AR(1) process in logs:

$$lnz_t = \psi lnz_{t-1} + e_{z,t} \tag{2.9}$$

with $\psi \in (0,1)$ indicating the persistence of the process, and $e_{z,t}$ normally distributed with mean zero and variance σ^2_z.

Let β^E denote the subjective discount factor for entrepreneurs. Following Iacoviello (2005), it is assumed that $\beta^E < \beta$ to ensure that in the steady state and its neighborhood, entrepreneurs are net borrowers and households are net savers. The necessity of this assumption is shown later after solving the entrepreneur's problem. The entrepreneur's objective is to maximize the expected lifetime utility:

$$E_t\sum_{s=0}^{\infty}(\beta^E)^s ln(c^E_{t+s}) \tag{2.10}$$

subject to a budget constraint (2.11) and a collateral constraint (2.12). Let $R_{b,t}$ denote the gross nominal loan rate in period t, then at the beginning of period $t+1$, the gross real loan interest payment is $\frac{R_{b,t}b_t}{\pi_{t+1}}$. Since the loan contract is denominated in nominal terms with a specified $R_{b,t}$, a rise in inflation in period $t+1$ reduces the firm's real debt burden. At the end of period t, entrepreneurs can sell the undepreciated capital $(1-\delta)k^E_{t-1}$ to capital producers at the real price of capital q_t, where $\delta \in (0,1)$ is the depreciation rate for physical capital. The wholesale good produced in period t is sold to retailers at a nominal price $p_{w,t}$. Let x_t denote the markup of the price of the final consumption good over the price of the wholesale good, that is, $x_t \equiv \frac{p_t}{p_{w,t}}$. In each period t, the outflow of funds due to consumption c^E_t, cost of capital investment $q_t[k^E_t - (1-\delta)k^E_{t-1}]$, cost of housing investment $q_{h,t}(h^E_t - h^E_{t-1})$, real wage payments to households $w_t l^E_t$ and real gross loan interest payments $\frac{R_{b,t-1}b_{t-1}}{\pi_t}$, would equal the inflow of funds due to the real revenue from selling the wholesale good $\frac{y_{w,t}}{x_t}$ and the loans granted by banks b_t. Hence the budget constraint in real terms is:

$$c_t^E + q_t k_t^E + q_{h,t} h_t^E + w_t l_t^E + \frac{R_{b,t-1}b_{t-1}}{\pi_t} = \frac{y_{w,t}}{x_t} + (1-\delta) q_t k_{t-1}^E + q_{h,t} h_{t-1}^E + b_t \tag{2.11}$$

An agency problem is introduced by assuming costly debt enforcement, based on Kiyotaki and Moore (1997). Assume the entrepreneurs face limited liability on debt obligations and if they repudiate their debt obligations, banks can only claim a fraction of their assets. Assuming both real estate and physical capital can be used as collateral assets,[6] let $m_h \in (0,1)$ and $m_k \in (0,1)$ denote the fractions of housing collateral and physical capital collateral respectively that can be confiscated by banks when the entrepreneurs fail to repay their debt. Consequently, the maximum amount an entrepreneur can borrow is such that the gross nominal debt interest payment $R_{b,t}b_t$ equals the expected value of assets that banks can claim after debt repudiation, which is equal to $m_{h,t}E_t\left[q_{h,t+1}h_t^E\pi_{t+1}\right] + m_{k,t}E_t\left[q_{t+1}k_t^E(1-\delta)\pi_{t+1}\right]$. As a result, the collateral constraint can be written as:

$$b_t \leqslant m_{h,t}E_t\left[\frac{q_{h,t+1}h_t^E\pi_{t+1}}{R_{b,t}}\right] + m_{k,t}E_t\left[\frac{q_{t+1}k_t^E(1-\delta)\pi_{t+1}}{R_{b,t}}\right] \tag{2.12}$$

The pledgeability ratios $m_{h,t}$ and $m_{k,t}$ are subject to the collateral shocks and follow an AR(1) process in logs:

$$lnm_{h,t} = (1-\psi_{m_h})lnm_h + \psi_{m_h}lnm_{h,t-1} + e_{m_h,t} \tag{2.13}$$

$$lnm_{k,t} = (1-\psi_{m_k})lnm_k + \psi_{m_k}lnm_{k,t-1} + e_{m_k,t} \tag{2.14}$$

where m_h and m_k are the steady state values, $\psi_{m_h} \in (0,1)$ and $\psi_{m_k} \in (0,1)$ indicate the persistence of the process, $e_{m_h,t}$ and $e_{m_k,t}$ are normally distributed with mean zero and variance $\sigma^2_{m_h}$ and $\sigma^2_{m_k}$, respectively. Since the pledgeability ratio resembles the loan-to-value ratio (i.e., amount of borrowing divided by the value of collateral), the collateral shock can also be interpreted as the macroprudential policy shock.

Let $\lambda_{1,t}^E$ and $\lambda_{2,t}^E$ denote the Lagrange multipliers associated with the budget constraint (2.11) and the borrowing constraint (2.12) respectively. Then the first order conditions with respect to the entrepreneur's consumption c_t^E (2.15), loan demand b_t (2.16), labor demand l_t^E (2.17), capital demand k_t^E (2.18) and housing demand h_t^E (2.19) are:

$$\frac{1}{c_t^E} = \lambda_{1,t}^E \tag{2.15}$$

[6] From this section onwards, assets refer to both real estate and physical capital. Although real estate is the only collateral in both Iacoviello (2005) and Andrés and Arce (2012), it is plausible to assume physical capital can also serve this purpose.

$$\lambda^E_{2,t} = \lambda^E_{1,t} - \beta^E E_t \left[\lambda^E_{1,t+1} \frac{R_{b,t}}{\pi_{t+1}} \right] \tag{2.16}$$

$$\frac{(1-\alpha-v)y_{w,t}}{x_t l^E_t} = w_t \tag{2.17}$$

$$q_t \lambda^E_{1,t} = \beta^E E_t \left[\lambda^E_{1,t+1} \left\{ \frac{\alpha y_{w,t+1}}{x_{t+1} k^E_t} + (1-\delta) q_{t+1} \right\} \right] + \lambda^E_{2,t} m_{k,t} E_t \left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}} \right] \tag{2.18}$$

$$q_{h,t} \lambda^E_{1,t} = \beta^E E_t \left[\lambda^E_{1,t+1} \left\{ \frac{v y_{w,t+1}}{x_{t+1} h^E_t} + q_{h,t+1} \right\} \right] + \lambda^E_{2,t} m_{h,t} E_t \left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}} \right] \tag{2.19}$$

Let variables without the time subscript denote the steady state values. Combining (2.15) and (2.16), it can be seen that in the steady state:

$$\lambda^E_2 = \frac{1}{c^E} \left(1 - \beta^E \frac{R_b}{\pi} \right) \tag{2.20}$$

From Euler equation (2.7) derived from the household's problem in Section 2.2.1, the steady state value of the gross real interest rate $\frac{R}{\pi}$ is determined by the household's subjective discount factor, such that $\frac{R}{\pi} = \frac{1}{\beta}$. Since $R_{b,t} = R_t$ under perfect banking competition, $\lambda^E_2 = \frac{1}{c^E}\left(1 - \frac{\beta^E}{\beta}\right)$. To ensure that the borrowing constraint always binds in the steady state, λ^E_2 must be positive, which implies $\beta^E < \beta$. This heterogeneity in the subjective discount factors guarantees that in the steady state, impatient entrepreneurs are net borrowers.[7]

Based on the budget constraint (2.11), define the entrepreneur's net worth n_t in period t after the productivity shock has been realized and output is produced, as the revenue accruing to the factor inputs of physical capital and real estate $\frac{(\alpha+v)y_{w,t}}{x_t}$, plus the total value of the real estate holdings and capital stock $q_{h,t}h^E_{t-1} + q_t(1-\delta)k^E_{t-1}$, and net of the gross real loan interest payment $\frac{R_{b,t-1}b_{t-1}}{\pi_t}$ at the beginning of period t.[8] Hence, n_t can be written as:

$$n_t \equiv \frac{(\alpha+v)y_{w,t}}{x_t} + q_t(1-\delta)k^E_{t-1} + q_{h,t}h^E_{t-1} - \frac{R_{b,t-1}b_{t-1}}{\pi_t} \tag{2.21}$$

[7]In this strand of the literature, it is a common approach to assume $\beta^E < \beta$ to ensure that the borrowing constraint always binds in the steady state and its neighborhood, as long as the size of the shocks are sufficiently small (Liu et al., 2013; Andrés and Arce, 2012; Gerali et al., 2010; Iacoviello, 2005). To ensure the borrowing constraint is always binding in this paper, the parameter restriction is imposed to guarantee a positive λ^E_2, and only adverse shocks are analysed, although the results are symmetric for positive shocks as long as the constraint remains binding.

[8]Theoretically, after a negative productivity shock, the output of the wholesale good $y_{w,t}$ can be low enough such that the entrepreneur is not able to repay the gross real loan interest repayment $\frac{R_{b,t-1}b_{t-1}}{\pi_t}$ and hence the net worth is negative. Under reasonable calibration, the steady state net worth is high, as shown in Table A.2 in Appendix A.4, such that even after a very large and persistent negative productivity shock (i.e., the size of the shock is five times the standard deviation and the persistence parameter is 0.97 in the AR(1) process for productivity), the net worth is far from being negative. Since the default on loans is extremely unlikely, this paper ignores the possibility of loan default. The possibility of entrepreneurs' default on loans and hence banks' default on liabilities (deposits) are studied in Chapter 3.

where $\frac{(\alpha+v)y_{w,t}}{x_t} = \frac{y_{w,t}}{x_t} - w_t l_t^E$, which follows from the first order condition with respect to l_t^E (2.17). Rewriting the budget constraint (2.11) in terms of n_t gives:

$$c_t^E + q_t k_t^E + q_{h,t} h_t^E = n_t + b_t \tag{2.22}$$

which implies the entrepreneur finances his consumption c_t^E and the purchase of new capital and housing $(q_t k_t^E + q_{h,t} h_t^E)$ by bank loans b_t and the retained earnings n_t. To derive c_t^E, use the first order conditions (2.15)-(2.19) and the binding borrowing constraint based on (2.12):

$$b_t = m_{h,t} E_t \left[\frac{q_{h,t+1} h_t^E \pi_{t+1}}{R_{b,t}} \right] + m_{k,t} E_t \left[\frac{q_{t+1} k_t^E (1-\delta) \pi_{t+1}}{R_{b,t}} \right] \tag{2.23}$$

It is proved in Appendix A.1 that due to the assumption of log utility, the entrepreneur's consumption in period t is a fixed proportion $(1-\beta^E)$ of net worth n_t:

$$c_t^E = (1-\beta^E) n_t \tag{2.24}$$

It follows from (2.22) that in the presence of a binding constraint, the real loan demand b_t is the total purchasing cost of new capital and housing in excess of the internal financing or savings $\beta^E n_t$:

$$b_t = q_t k_t^E + q_{h,t} h_t^E - \beta^E n_t \tag{2.25}$$

where $\beta^E n_t$ is the part of net worth that is not consumed and can thus be used to finance the purchase of physical capital and housing. Note that (2.23) effectively determines market loan demand due to the binding borrowing constraint and it implies an inverse relation between the equilibrium loan rate $R_{b,t}$ and loan quantity b_t.[9] In particular, for given (asset) prices $q_{h,t+1}$, q_{t+1} and π_{t+1}, a higher loan rate $R_{b,t}$ corresponds to a lower loan quantity b_t, both directly and indirectly through its effect on the entrepreneur's housing and capital demand, h_t^E and k_t^E. It is shown in Appendix A.3.1 that $\frac{\partial h_t^E}{\partial R_{b,t}} < 0$ and $\frac{\partial k_t^E}{\partial R_{b,t}} < 0$.

Capital Producers

Perfectly competitive capital producers are introduced to derive an explicit expression for the real price of capital q_t (Gambacorta and Signoretti, 2014).[10] They purchase undepreciated capital $(1-\delta)k_{t-1}$ at the real price q_t from entrepreneurs and i_t units of

[9] With perfect banking competition, $R_{b,t}$ is given by the gross deposit rate R_t, so b_t is determined. With imperfect banking competition, each individual bank takes into account the effect of its choice on b_t and hence $R_{b,t}$.

[10] In a standard RBC model, the price of physical capital relative to consumption is one. q_t is an important variable here because it can affect the entrepreneur's collateral value and net worth.

final consumption goods from retailers to produce new capital k_t at the end of period t:

$$k_t = i_t + (1-\delta)k_{t-1} \tag{2.26}$$

where i_t is also gross investment. The new capital produced will be sold back to the entrepreneur at the real price q_t, which will be used in the production of the wholesale good in period $t+1$. Following Christiano et al. (2005), assume capital producers face investment adjustment costs that depend on the gross growth rate of investment $\frac{i_t}{i_{t-1}}$. Assume old capital can be converted one-to-one into new capital and a quadratic unit investment adjustment cost $f\left(\frac{i_t}{i_{t-1}}\right) = \frac{\chi}{2}\left(\frac{i_t}{i_{t-1}} - 1\right)^2$ is only incurred in the production of new capital when using the final consumption good as the input, where $f(1) = f'(1) = 0$, $f''(1) > 0$ and $\chi > 0$. This specification of the adjustment cost implies that fewer units of new capital would be produced from one unit of investment whenever $\frac{i_t}{i_{t-1}}$ deviates from its steady state value of one and the parameter χ reflects the magnitude of the cost.

Hence, the representative capital producer chooses the gross investment level i_t to maximize the sum of the expected discounted future profits made from the sales revenue of new capital $q_t k_t$ net of the input cost $[q_t(1-\delta)k_{t-1} + i_t]$ and the investment adjustment cost $f\left(\frac{i_t}{i_{t-1}}\right) i_t$:

$$E_t \sum_{s=0}^{\infty} \Lambda_{t,t+s} \left[q_t k_t - q_t(1-\delta)k_{t-1} - i_t - \frac{\chi}{2}\left(\frac{i_t}{i_{t-1}} - 1\right)^2 i_t \right] \tag{2.27}$$

where $\Lambda_{t,t+s} \equiv \beta^s \frac{u'(c_{t+s})}{u'(c_t)}$ is the stochastic discount factor, since households own the capital producers. Using (2.26), the objective function (2.27) can be simplified to:

$$E_t \sum_{s=0}^{\infty} \Lambda_{t,t+s} \left[(q_t - 1)i_t - \frac{\chi}{2}\left(\frac{i_t}{i_{t-1}} - 1\right)^2 i_t \right] \tag{2.28}$$

Taking the first order condition with respect to investment i_t gives the following expression for the real price of capital:

$$q_t = 1 + \frac{\chi}{2}\left(\frac{i_t}{i_{t-1}} - 1\right)^2 + \chi \frac{i_t}{i_{t-1}}\left(\frac{i_t}{i_{t-1}} - 1\right) - \chi E_t \left[\Lambda_{t,t+1} \left(\frac{i_{t+1}}{i_t}\right)^2 \left(\frac{i_{t+1}}{i_t} - 1\right) \right] \tag{2.29}$$

In the steady state, the real price of capital q is one, since $i_{t+1} = i_t = i_{t-1}$. Any real profits Π_t^{CP} (which only arise outside the steady state) are rebated to the households, where $\Pi_t^{CP} = (q_t - 1)i_t - \frac{\chi}{2}\left(\frac{i_t}{i_{t-1}} - 1\right)^2 i_t$.

Retailers

Following Bernanke et al. (1999), retailers are assumed to be monopolistically competitive. A continuum of retailers of unit mass, indexed by j, buy the wholesale good at a nominal price $p_{w,t}$ from entrepreneurs and use it as the only input to produce differentiated retail goods costlessly. Assume that one unit of the wholesale good can produce one unit of the differentiated product, so the marginal cost of production is the real price of the wholesale good $\frac{p_{w,t}}{p_t}$. Each retailer j produces a different variety $y_t(j)$ and charges a nominal price $p_t(j)$ for the differentiated product. The output of the final consumption good y_t is a constant elasticity of substitution (CES) composite of all the different varieties produced by the retailers, using the Dixit and Stiglitz (1977) framework:

$$y_t = \left[\int_0^1 y_t(j)^{\frac{\epsilon-1}{\epsilon}} dj\right]^{\frac{\epsilon}{\epsilon-1}} \tag{2.30}$$

where $\epsilon > 1$ is the elasticity of intratemporal substitution between different varieties. Given the aggregate output index y_t, it can be calculated from the cost minimization problem of the buyers of the final consumption good that each retailer j faces a downward-sloping demand curve:

$$y_t(j) = \left[\frac{p_t(j)}{p_t}\right]^{-\epsilon} y_t \tag{2.31}$$

It can be shown that the aggregate consumption-based price index is:

$$p_t = \left[\int_0^1 p_t(j)^{1-\epsilon} dj\right]^{\frac{1}{1-\epsilon}} \tag{2.32}$$

which is defined as the minimum expenditure to obtain one unit of consumption y_t in the cost minimization problem for the final output users.

Under monopolistic competition, retailers have price setting power, which is essential for introducing the nominal price rigidity *à la* Calvo (1983). With a nominal rigidity, monetary policy has real effects and the impact of a monetary policy shock can be analysed. Each retailer j sets its own price $p_t(j)$ taking the aggregate price p_t and the demand curve (2.31) as given. Under Calvo pricing, each retailer j is only allowed to change its price $p_t(j)$ in period t with probability $(1-\theta)$. The probability of price adjustment is independent of the time since the last adjustment, so in each period, a fraction $(1-\theta)$ of retailers reset their prices, whereas a fraction θ of retailers keep their prices fixed. Hence, $\theta \in (0,1)$ reflects the degree of price stickiness. Let $p_t^*(j)$ denote the optimal reset price in period t, then the corresponding demand facing retailer j who adjusted its price in period t, but cannot adjust its price in period $t+s$, is:

$$y_{t+s}^*(j) = \left[\frac{p_t^*(j)}{p_{t+s}}\right]^{-\epsilon} y_{t+s} \tag{2.33}$$

Retailer j chooses $p_t^*(j)$ to maximize the expected discounted value of real profits while its price is kept fixed at $p_t^*(j)$:

$$\sum_{s=0}^{\infty} \theta^s E_t \left[\Lambda_{t,t+s} \left\{\frac{p_t^*(j)}{p_{t+s}} y_{t+s}^*(j) - \frac{1}{x_{t+s}} y_{t+s}^*(j)\right\}\right] \tag{2.34}$$

subject to the demand function (2.33), where $\Lambda_{t,t+s} \equiv \beta^s \frac{u'(c_{t+s})}{u'(c_t)}$ is the stochastic discount factor, since households own the retailers, θ^s is the probability that $p_t^*(j)$ would remain fixed for s periods, and $\frac{1}{x_{t+s}} = \frac{p_{w,t+s}}{p_{t+s}}$ is the price of the wholesale good in terms of the consumption units or the real marginal cost of production in period $t+s$. Taking the first order condition to solve for $p_t^*(j)$ gives the following optimal pricing equation:

$$p_t^*(j) = \frac{\epsilon}{\epsilon - 1} \frac{\sum_{s=0}^{\infty} (\beta\theta)^s E_t \left[u'(c_{t+s}) x_{t+s}^{-1} p_{t+s}^{\epsilon} y_{t+s} \right]}{\sum_{s=0}^{\infty} (\beta\theta)^s E_t \left[u'(c_{t+s}) p_{t+s}^{\epsilon-1} y_{t+s} \right]} \tag{2.35}$$

The derivation is shown in Appendix A.2.1. In a symmetric equilibrium, all the retailers that adjust their prices in period t will set the same optimal price, such that $p_t^*(j) = p_t^*$. It is proved in Appendix A.2.2 that the aggregate price level evolves as follows:

$$p_t^{1-\epsilon} = \theta p_{t-1}^{1-\epsilon} + (1-\theta)(p_t^*)^{1-\epsilon} \tag{2.36}$$

which is independent of the heterogeneity of the retailers due to the convenience of the Calvo assumption. With randomly chosen price-adjusting retailers and the large number of retailers, there is no need to keep track of each retailer's price evolution.

Since there is a one-to-one conversion rate from the wholesale good to the differentiated retail good, in equilibrium the supply of wholesale good output $y_{w,t}$ is equal to the demand $y_t(j)$ over the entire unit interval of retailers j. Using retailer $j's$ individual demand function (2.31), the wholesale good output can be expressed as:

$$y_{w,t} = \int_0^1 y_t(j) dj = y_t \int_0^1 \left[\frac{p_t(j)}{p_t} \right]^{-\epsilon} dj \tag{2.37}$$

As seen from the above equation, the final consumption good output y_t differs from the wholesale good output $y_{w,t}$ by a factor of the price dispersion $\int_0^1 \left[\frac{p_t(j)}{p_t} \right]^{-\epsilon} dj$. In a zero-inflation steady state, the price dispersion is one and the final output y_t would equal the wholesale good output $y_{w,t}$. Use (2.37) and let $f_{3,t} \equiv \int_0^1 \left[\frac{p_t(j)}{p_t} \right]^{-\epsilon} dj$ denote the price dispersion, then the real profit Π_t^R made by the retailers is:

$$\Pi_t^R = y_t - \frac{y_{w,t}}{x_t} = \left(\frac{1}{f_{3,t}} - \frac{1}{x_t} \right) y_{w,t} \tag{2.38}$$

which will be rebated lump sum back to the households. The recursive formulation of the price dispersion used for numerical computation and the derivation for Π_t^R are shown in Appendix A.2.3.

Banking Sector

Assume there is a continuum of banks of mass one, indexed by j, which are perfectly competitive with no price-setting power. The gross nominal interest rate R_t is controlled by the central bank and is thus taken as given. Following Andrés and Arce (2012), assume all bank profits $\Pi_t^B(j)$ are distributed as dividends to households each period, so $\Pi_t^B = \sum_j \Pi_t^B(j)$. In addition, assume there is zero bank capital, so bank loans (assets) equal the deposits (liabilities):

$$b_t(j) = d_t(j) \tag{2.39}$$

In each period t, the total outflow of funds, consisting of the dividend payment to households $\Pi_t^B(j)$, loans granted to firms $b_t(j)$, and the gross real deposit interest payments to households $\frac{R_{t-1}d_{t-1}(j)}{\pi_t}$, equals the total inflow of funds from the deposits saved by households $d_t(j)$ and the gross real loan interest payments received from firms $\frac{R_{b,t-1}b_{t-1}(j)}{\pi_t}$. Assuming costless financial intermediation and no default on loans,[11] each bank j faces the following budget constraint:

$$\Pi_t^B(j) + b_t(j) + \frac{R_{t-1}d_{t-1}(j)}{\pi_t} = d_t(j) + \frac{R_{b,t-1}b_{t-1}(j)}{\pi_t} \tag{2.40}$$

Each bank j chooses the units of loans $b_t(j)$ and the units of deposits $d_t(j)$ to maximize the sum of the expected discounted value of real profits:

$$E_t \sum_{s=0}^{\infty} \Lambda_{t,t+s} \Pi_t^B(j) \tag{2.41}$$

subject to the balance sheet identity (2.39) and the budget constraint in real terms (2.40). Substituting (2.39) into (2.40) simplifies the bank's real profit $\Pi_t^B(j)$ to:

$$\Pi_t^B(j) = \frac{1}{\pi_t}(R_{b,t-1} - R_{t-1})b_{t-1}(j) \tag{2.42}$$

Taking the first order condition of (2.41) with respect to $b_t(j)$ gives:

$$E_t \left[\Lambda_{t,t+1} \frac{1}{\pi_{t+1}} (R_{b,t} - R_t) \right] = 0 \tag{2.43}$$

Since $\Lambda_{t,t+1} > 0$ and $\pi_{t+1} \equiv \frac{p_{t+1}}{p_t} > 0$, the nominal loan interest margin $(R_{b,t} - R_t)$ is zero. With perfect banking competition and no expected default on loans, the market-determined gross nominal loan rate $R_{b,t}$ equals R_t.

Central Bank

Suppose monetary policy is implemented by a Taylor rule with interest rate smoothing, which responds to both the deviation of the gross inflation rate from the inflation target π and the deviation of output from its steady state y. The central bank controls the gross nominal interest rate R_t on risk-free bonds and bank deposits, following the Taylor

[11] Recall that this paper neglects the possibility of loan default since it is extremely unlikely under reasonable calibration, as discussed in Section 2.2.1.

rule specification below:

$$R_t = \rho_r R_{t-1} + (1-\rho_r)\left[R + \kappa_\pi(\pi_t - \pi) + \kappa_y\left(y_t - y\right)\right] + e_{r,t} \tag{2.44}$$

where variables without time subscript represent steady state values, and $e_{r,t}$ is a monetary policy shock which is a white noise process with zero mean and variance σ_r^2. The coefficient $\rho_r \in (0,1)$ is the interest rate smoothing parameter, and κ_π and κ_y are non-negative feedback parameters that reflect the sensitivity of the interest rate to output and inflation deviations. In the baseline analysis, I set κ_y to be zero for simplicity.[12] Due to interest rate smoothing, this policy rule implies a partial adjustment of R_t. As can be seen from (2.44), R_t is a weighted average of the lagged nominal interest rate R_{t-1} and the current target rate, which depends positively on the deviation of inflation from its target and the deviation of output from its steady state value. Let $R_{r,t}$ denote the gross real interest rate, then the relation between the nominal and real interest rates is given by the Fisher equation:

$$R_{r,t} = E_t\left[\frac{R_t}{\pi_{t+1}}\right] \tag{2.45}$$

2.2.2 Imperfect Banking Competition

Imperfect banking competition is analysed by replacing the perfectly competitive banking sector by a Cournot banking sector. As the banking sector tends to be dominated by a few large players, a Cournot banking sector is used to characterise oligopolistic competition. In a Cournot equilibrium, banks' quantity-setting decisions affect the market loan rate. The model set-up is unchanged apart from the banking sector which is now imperfectly competitive. Only the differences from Section 2.2.1 are discussed here. Assume now there are N banks in the economy, indexed by j, which operate under Cournot competition. Each individual bank takes the quantities of loans chosen by the other banks $m \neq j$ as given. However, it takes into account the effect of its choice $b_t(j)$ on the (partial) equilibrium in the loan market, through the total loan quantity b_t and the loan rate $R_{b,t}$, but it ignores general equilibrium effects and takes other prices and aggregate quantities as given.[13] Each bank j sets the quantity of loans $b_t(j)$ to maximize the sum of the present discounted value of future profits:

$$E_t \sum_{s=0}^{\infty} \Lambda_{t,t+s} \Pi_t^B(j) \tag{2.46}$$

[12]When $\kappa_y = 0$, the Taylor principle implies that $\kappa_\pi > 1$ will ensure the nominal interest rate R_t is raised sufficiently in response to an increase in the gross inflation rate π_t so that the real interest rate rises.

[13]Using (2.23), the effect of $R_{b,t}$ on b_t includes the effect through capital k_t^E and housing h_t^E, taking the inflation rate π_{t+1} and the real prices of housing $q_{h,t+1}$ and capital q_{t+1} as given.

where

$$\Pi_t^B(j) = \frac{1}{\pi_t}\left[R_{b,t-1}\left(b_{t-1}(j) + \sum_{m\neq j} b_{t-1}(m)\right) - R_{t-1}\right] b_{t-1}(j) \tag{2.47}$$

The real profit $\Pi_t^B(j)$ is positive due to imperfect competition and will be rebated back to the households. A key difference from Section 2.2.1 is that $R_{b,t}(.)$ now represents the inverse loan demand function, which depends on b_t and thereby $b_t(j)$. This is crucial for introducing imperfect banking competition. The dependence of $R_{b,t}$ on $b_t(j)$ means that each bank j has some control over the equilibrium gross loan interest rate by altering its own quantity of loans given the other banks' loan quantities and this is taken into consideration by bank j under Cournot competition when choosing $b_t(j)$. Solving the profit maximization problem with respect to $b_t(j)$ gives the following first order condition:

$$E_t\left[\Lambda_{t,t+1}\frac{1}{\pi_{t+1}}\left\{\frac{\partial R_{b,t}}{\partial b_t(j)}b_t(j) + R_{b,t} - R_t\right\}\right] = 0 \tag{2.48}$$

In a Cournot equilibrium, the total optimal loan quantity is $b_t = b_t(j) + \sum_{m\neq j} b_t(m)$ and each bank produces a share of the total quantity. Assuming banks are identical, then $b_t(j) = \frac{b_t}{N}$ in equilibrium. Since $\frac{\partial R_{b,t}}{\partial b_t(j)} = \frac{\partial R_{b,t}}{\partial b_t}\frac{\partial b_t}{\partial b_t(j)} = \frac{\partial R_{b,t}}{\partial b_t}$, in Cournot equilibrium, the first order condition (2.48) can be rewritten as:

$$E_t\left[\Lambda_{t,t+1}\frac{1}{\pi_{t+1}}\left\{\frac{\partial R_{b,t}}{\partial b_t}\frac{b_t}{N} + R_{b,t} - R_t\right\}\right] = 0 \tag{2.49}$$

where the market loan demand b_t and the partial derivative $\frac{\partial R_{b,t}}{\partial b_t}$ can be calculated explicitly from the representative entrepreneur's problem in Section 2.2.1.[14] The market loan demand is given by the binding collateral constraint (2.23). As can be seen from (2.23), $R_{b,t}$ has a direct negative effect on market loan demand b_t since an increase in $R_{b,t}$ reduces the entrepreneur's borrowing capacity. Besides, $R_{b,t}$ also has an indirect effect on b_t by influencing the entrepreneur's demand for housing and physical capital, which can be seen from the first order conditions for housing (2.19) and physical capital (2.18). Hence, when bank j chooses $b_t(j)$, which affects the equilibrium gross loan rate $R_{b,t}$ under Cournot competition, it needs to consider how entrepreneurs would respond by changing their demand for physical capital $\frac{\partial k_t^E}{\partial R_{b,t}}$ and housing $\frac{\partial h_t^E}{\partial R_{b,t}}$. Taking the total derivative of b_t in (2.23) with respect to $R_{b,t}$ and using (2.23) to simplify:

$$\frac{\partial b_t}{\partial R_{b,t}} = -\frac{b_t}{R_{b,t}} + m_{h,t}E_t\left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}}\right]\frac{\partial h_t^E}{\partial R_{b,t}} + m_{k,t}E_t\left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}}\right]\frac{\partial k_t^E}{\partial R_{b,t}} < 0 \tag{2.50}$$

It is shown in Appendix A.3.1 that the partial derivatives $\frac{\partial h_t^E}{\partial R_{b,t}} < 0$ and $\frac{\partial k_t^E}{\partial R_{b,t}} < 0$ can be calculated from the first order conditions (2.19) and (2.18) and have negative signs, which

[14] In equilibrium, the total supply of loans from the Cournot banking sector equals the market loan demand (2.23) from the entrepreneur's problem.

means an increase in $R_{b,t}$ reduces the entrepreneur's demand for housing and physical capital. In addition, Appendix A.3.1 shows that these partial derivatives ($\frac{\partial h_t^E}{\partial R_{b,t}}$ and $\frac{\partial k_t^E}{\partial R_{b,t}}$) only depend on period t variables and expectations, and therefore so do $\frac{\partial b_t}{\partial R_{b,t}}$ using (2.50) and the loan demand elasticity $PED_t \equiv -\frac{\partial b_t}{\partial R_{b,t}} \frac{R_{b,t}}{b_t}$. Since $\Lambda_{t,t+1} > 0$ and $\pi_{t+1} \equiv \frac{p_{t+1}}{p_t} > 0$, (2.49) implies that $\left(\frac{\partial R_{b,t}}{\partial b_t} \frac{b_t}{R_{b,t}} \frac{1}{N} + 1\right) R_{b,t} = R_t$. It follows that the equilibrium loan rate depends on the policy rate R_t, the number of banks N, and the price elasticity of loan demand $PED_t \equiv -\frac{\partial b_t}{\partial R_{b,t}} \frac{R_{b,t}}{b_t} > 0$:

$$R_{b,t} = \frac{1}{1 - \frac{1}{N} PED_t^{-1}} R_t \tag{2.51}$$

As can be seen, a larger number of banks N (implying more intense banking competition) and/or a higher loan demand elasticity PED_t lead to a lower loan rate. With perfect banking competition, each bank faces a perfectly elastic loan demand, so $NPED_t \to \infty$ and $R_{b,t} = R_t$, although the market loan demand given by (2.23) is downward-sloping, as shown in Appendix A.3.2. With Cournot competition, banks with market power can affect the equilibrium loan rate by taking advantage of the endogenously changing loan demand elasticity. After an adverse shock, the entrepreneur is more financially constrained and hence the loan demand becomes more inelastic (i.e., PED_t falls), leading to a higher loan rate and higher loan interest margin ($R_{b,t} - R_t$) for a given level of R_t. This countercyclical loan interest margin can amplify aggregate fluctuations.

As the change in loan demand elasticity determines how the loan margin responds to shocks, it is important to understand what determines the elasticity. It is shown in Appendix A.3.2 that the price elasticity of the market loan demand PED_t decreases (more inelastic loan demand) when the expected marginal products of capital and housing are higher and/or the expected future prices of capital and housing are lower. This is because in the presence of a binding collateral constraint, higher expected marginal products imply that the entrepreneur is more financially constrained and cannot finance the purchase of the optimal amount of production inputs (capital and housing). Lower expected asset prices directly reduce the entrepreneur's borrowing capacity through the binding constraint (2.23). Hence, both higher expected marginal products and lower expected asset prices imply that the entrepreneur is more financially constrained and hence the loan demand is more inelastic. Besides, PED_t is also decreasing in the pledgeability ratios $m_{h,t}$ and $m_{k,t}$ (e.g., after a negative collateral shock), as shown in Appendix A.3.2. A lower pledgeability ratio directly reduces the entrepreneur's borrowing capacity, leading to a more inelastic loan demand. A lower borrowing capacity also raises the expected marginal products of capital and housing and reduces the expected asset prices due to a lower demand for capital and housing, thus further reducing the loan demand elasticity.

2.2.3 Equilibrium Conditions

In equilibrium, the aggregate resource constraint is:

$$c_t + c_t^E + i_t + \frac{\chi}{2}\left(\frac{i_t}{i_{t-1}} - 1\right)^2 i_t = y_t \tag{2.52}$$

which is also the goods market clearing condition. Let b_t^B and d_t^B denote the total units of loans given out and deposits taken in by the banking sector, respectively. Under perfect banking competition with a continuum of banks of unit mass, $b_t^B = \int_0^1 b_t(j)dj$ and $d_t^B = \int_0^1 d_t(j)dj$, while with a Cournot banking sector, $b_t^B = \sum_{j=1}^N b_t(j)$ and $d_t^B = \sum_{j=1}^N d_t(j)$ in equilibrium. The other market clearing conditions are: (labor) $l_t = l_t^E$, (capital) $k_t = k_t^E$, (housing) $h_t + h_t^E = 1$, where the total fixed supply of housing is normalised to one, (deposits) $d_t = d_t^B$, (loans) $b_t = b_t^B$, and (financial intermediation) $b_t^B = d_t^B$.

2.3 Calibration

The models with two types of banking competition are solved numerically using Dynare after calibrating the parameters to a quarterly frequency. Using the ECB's harmonized monetary financial institutions' (MFI) interest rates from 2000 to 2018, the average annualised household deposit rate is around 2.16%. Hence, the household subjective discount factor β, is set at 0.995, giving an annualised net real deposit rate of $\left(\frac{1}{0.995} - 1\right) * 4 \approx 2\%$. The subjective discount factor for the entrepreneur $\beta^E = 0.97$ is taken from Andrés and Arce (2012). As shown in Section 2.2.1, the subjective discount factor for entrepreneurs β^E needs to be smaller than β, to ensure a binding collateral constraint in the steady state. When imperfect banking competition is introduced into the framework of collateral constraints, the restriction to ensure a binding borrowing constraint is no longer $\beta^E < \beta$, because the loan interest margin is greater than zero. Since $\lambda_2^E = \frac{1}{c^E}\left(1 - \beta^E \frac{R_b}{\pi}\right)$ and $\frac{R}{\pi} = \frac{1}{\beta}$, as shown by equations (2.20) and (2.7), $\lambda_2^E = \frac{1}{c^E}\left(1 - \frac{\beta^E}{\beta}\frac{R_b}{R}\right)$. Hence, as long as $\frac{R_b}{R} < \frac{\beta}{\beta^E}$ under imperfect banking competition, λ_2^E will be positive, which means the borrowing constraint will bind in the steady state (Andrés and Arce, 2012). As a result, when replacing the perfectly competitive banking sector with the Cournot banking sector, the ratio of gross nominal loan rate to gross nominal deposit rate in the steady state $\frac{R_b}{R}$ must satisfy the following condition:

$$1 \leqslant \frac{R_b}{R} < \frac{\beta}{\beta^E} \tag{2.53}$$

where the first inequality comes from the nonnegativity of bank profits. The upper bound $\frac{\beta}{\beta^E}$ imposes a limit on the size of the loan margin, which is around 10 percent points for an annualised loan interest margin under the calibration for β and β^E.

Standard parameters such as the physical capital share α, the depreciation rate δ, and the elasticity of substitution among differentiated retail goods ϵ are chosen to be 0.33, 0.025, and 6, respectively. The gross inflation target π is set at 1. In this zero-inflation steady state, $\epsilon = 6$ implies a final good price markup over the wholesale good of 20% since $x = \frac{\epsilon}{\epsilon - 1}$. Given the above parameters, the relative utility weight on leisure time ϕ_l is set at 1.45 to achieve steady state labor hours l of around 0.33. The relative utility weight on the holdings of real estate for households, $\phi_h = 0.1$, and the real estate share of the wholesale good output, $v = 0.05$, are taken from Andrés and Arce (2012) to achieve $h^E = 0.22$ under perfect banking competition in the steady state. The investment adjustment cost parameter χ does not affect the steady state and is set at 10, which is in line with Gerali et al. (2010). A larger χ reduces the magnitude of the responses of investment and physical capital after shocks.[15]

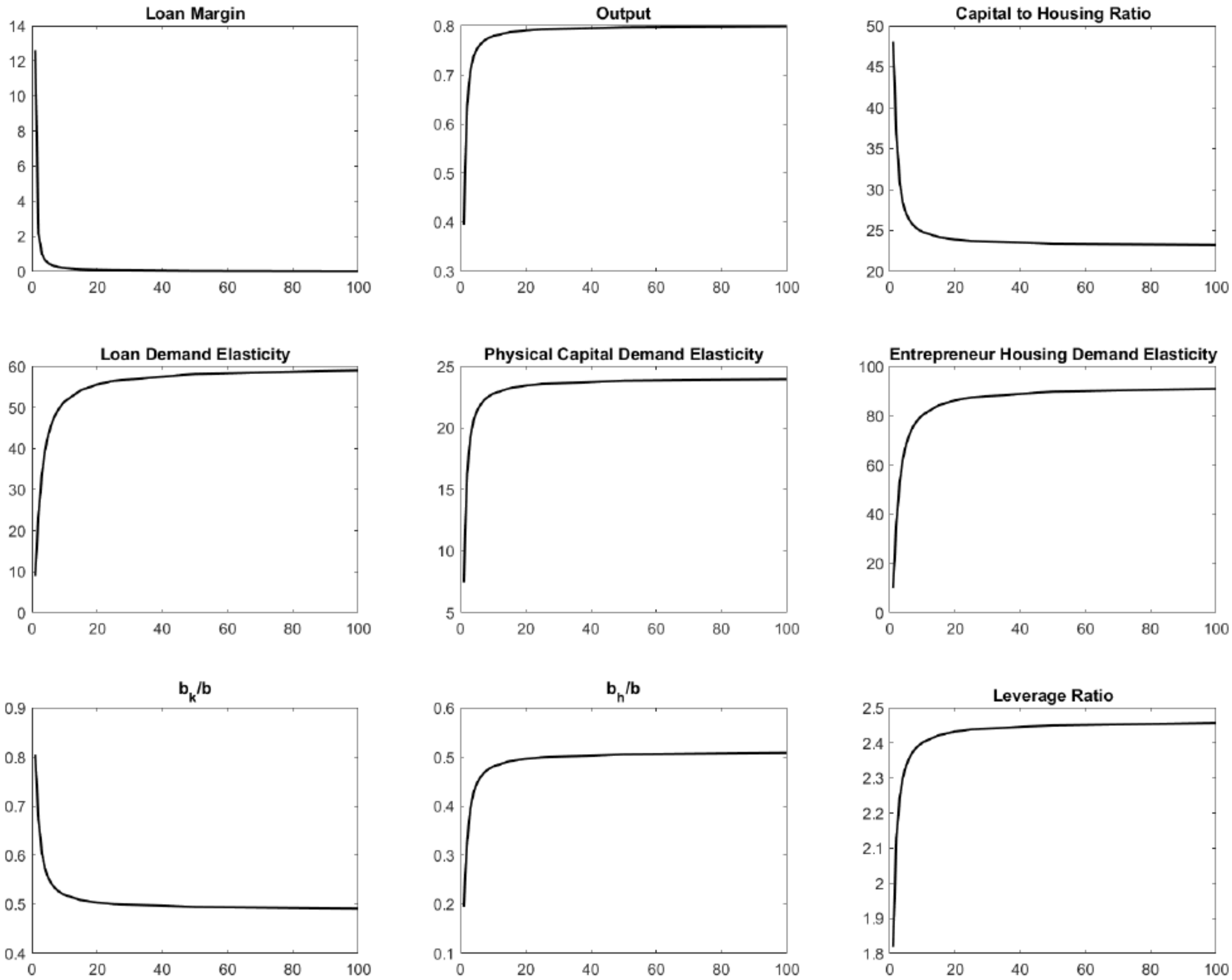
The probability θ of retailers keeping prices fixed in each period, the parameters in the Taylor rule and the shock-related parameters are in line with the literature (e.g., Andrés and Arce, 2012; Gerali et al., 2010). The probability θ is set at 0.75 to give a price rigidity of $\frac{1}{1-0.75} = 4$ quarters on average. The interest rate smoothing parameter ρ_r and the feedback coefficient on inflation κ_π are specified to be 0.8 and 1.5, respectively. The feedback coefficient on output κ_y is set to zero in the baseline analysis to simplify the Taylor rule. I also use different calibrations for ρ_r, κ_π and κ_y as robustness checks. The persistence parameters of the productivity shock ψ, the loan to value ratio shocks ψ_{m_h} and ψ_{m_k} are 0.97, 0.8 and 0.8, respectively. Besides, the standard deviation of the monetary policy shock σ_r is 0.001 and the standard deviations of the productivity shock σ_z, and the collateral shocks σ_{m_h} and σ_{m_k} are all set to be 0.01.

The steady-state loan-to-value ratios for housing and physical capital are chosen to be $m_h = 0.8$ and $m_k = 0.5$ respectively. The calibration for m_h is within the range set in the literature (e.g., Andrés and Arce, 2012; Gerali et al., 2010; Iacoviello, 2005). Note that m_k is set at a smaller value than m_h because physical capital often has a lower resale value than real estate. Given $m_h = 0.8$, $m_k = 0.5$, $\beta = 0.995$, $\beta^E = 0.97$, $\alpha = 0.33$, $v = 0.05$ and $\delta = 0.025$, the number of banks N is set to 4 to get a steady-state gross loan rate R_b of 1.01, implying an annualised net real loan rate of $(1.01 - 1) * 4 \approx 4\%$ and a real loan margin of around 200 basis points. This matches the average annualised corporate loan rate of around 4.14% and the loan interest margin of around 198 basis points across EU countries over the past 19 years, using the ECB's harmonized monetary financial institutions' (MFI) interest rates from 2000 to 2018. The annualised loan interest margin of around 2 percent points satisfies the condition $\frac{R_b}{R} \in [1, \frac{\beta}{\beta^E})$, and hence the borrowing constraint is binding in the steady state with a Cournot banking sector. A summary of the calibrated parameters is shown in Table A.1 in Appendix A.4. Given the calibration

[15] This is discussed in Section 2.5 and the impulse responses under different calibrations for χ are shown in Figure A.8 in Appendix A.6.

in this Section, the steady state values of the key variables are summarised in Table A.2 in Appendix A.4.

Figure 2.2: Steady State Values for Different Number of Banks



Note: The figure shows the steady state values of variables, with the loan margin expressed in percent points, for different number of banks N ranging from 1 to 100. The calibration for the other parameters can be found in Table A.1 in Appendix A.4. The second row plots the elasticities of loan demand, capital demand and housing demand to the loan rate, for different values of N. The leverage ratio is computed as the value of total assets $(qk^E + q_h h^E)$ over net worth n. The partial leverage ratios for capital $\frac{b_k}{b} \equiv m_k \frac{qk^E(1-\delta)\pi}{R_b b}$ and housing $\frac{b_h}{b} \equiv m_h \frac{q_h h^E(1-\delta)\pi}{R_b b}$ reflect the fractions of total borrowing against the capital and housing collateral, respectively, where $b_k + b_h = 1$.

Figure 2.2 plots the steady state values for some variables against the number of banks N that ranges from 1 to 100. A higher N implies more intense competition. When there is a monopoly bank, the annualised loan margin is around 1200*4 = 4800 basis points. Only in this case will the borrowing constraint not be binding. As can be seen, as N increases, the loan margin $(R_b - R)$ approaches zero and output increases. Besides, a lower loan rate raises the borrowing capacity of the entrepreneur through the binding collateral constraint (2.23). As a result, the leverage ratio rises and as the entrepreneur is less financially constrained, the loan demand elasticity PED is also higher.

One interesting aspect to notice is that the steady state capital-to-housing ratio falls when the loan rate declines as the number of banks N rises. Using the first order conditions (2.18) and (2.19), the steady state capital-to-housing ratio is:

$$\frac{k^E}{h^E} = \frac{\alpha q_h \left[1 - \beta^E - m_h \left(\frac{\pi}{R_b} - \beta^E\right)\right]}{vq\left[1 - \beta^E(1-\delta) - m_k(1-\delta)\left(\frac{\pi}{R_b} - \beta^E\right)\right]} = \frac{\alpha R_h}{vR_k} \tag{2.54}$$

where $R_h \equiv \frac{1}{\beta^E} q_h \left[1 - \beta^E - m_h \left(\frac{\pi}{R_b} - \beta^E\right)\right]$ equals the marginal product of housing $\frac{vy_w}{xh^E}$ and $R_k \equiv \frac{1}{\beta^E} q \left[1 - \beta^E(1-\delta) - m_k(1-\delta)\left(\frac{\pi}{R_b} - \beta^E\right)\right]$ equals the marginal product of capital $\frac{\alpha y_w}{xk}$ in terms of the final good. If the collateral constraint were nonbinding in the steady state, λ_2^E in (2.20) would equal zero and thereby $\left(\frac{\pi}{R_b} - \beta^E\right) = 0$, resulting in a capital-to-housing ratio $\frac{k^E}{h^E} = \frac{\alpha q_h\left(1-\beta^E\right)}{vq[1-\beta^E(1-\delta)]}$. But when the collateral constraint is binding, as for the calibration in this paper, it can be shown that the capital-to-housing ratio falls as the loan rate decreases if $\frac{R_k}{R_h} > \frac{m_k q(1-\delta)}{m_h q_h}$. This condition is satisfied if m_h is sufficiently large compared to m_k because (2.54) shows that a higher m_h relative to m_k reduces the user cost of housing R_h relative to that of capital R_k. Intuitively, for a given reduction in loan rate as banking competition becomes more intense, a higher h^E relaxes the binding collateral constraint (2.23) more than capital as $m_h > m_k$. As a result, the entrepreneur devotes more of the additional borrowing to housing, and the capital-to-housing ratio falls as the loan rate declines. This is also reflected by the higher elasticity of the entrepreneur's housing demand to the loan rate, as shown in Figure 2.2.

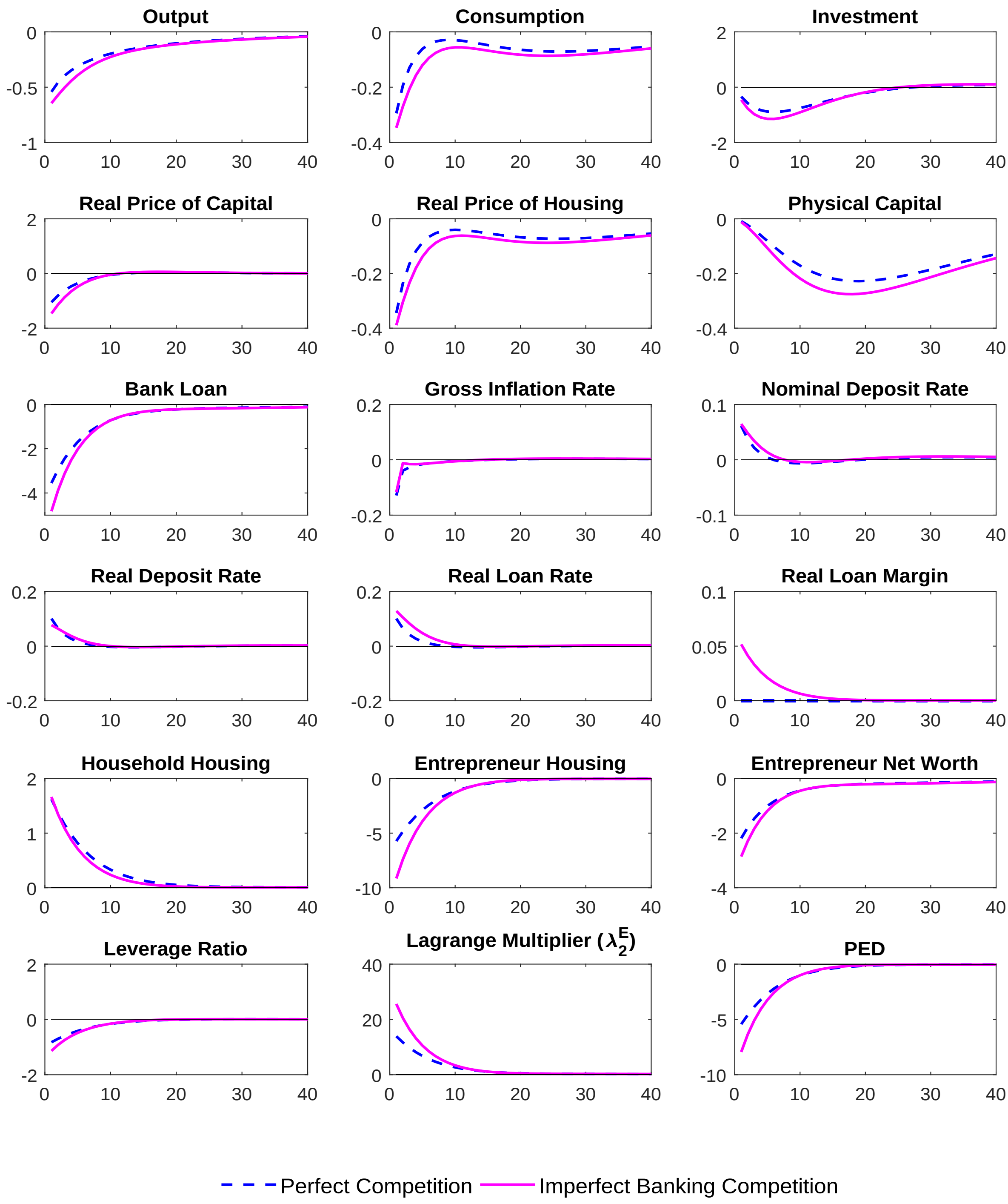
2.4 Dynamic Analysis

The impulse responses of some key variables under the two types of banking competition are compared after a contractionary monetary policy shock, a negative productivity shock and collateral or macroprudential policy shocks to the pledgeability ratios m_k and m_h. The model is solved numerically using Dynare, in which the impulse response functions (IRFs) are computed as the deviation in the trajectory of a variable from its steady state value following a shock at the beginning of period 1. This nonlinear model is solved using a first-order Taylor approximation around the steady state and the responses of variables are expressed as the percentage deviations from the steady state. All the equations used to compute the model are shown in Appendix A.5.

2.4.1 Monetary Policy Shock

An unexpected one-time monetary policy shock is implemented by a one standard deviation ($\sigma_r = 0.001$) or 10 basis points increase in the white noise term $e_{r,t}$ in the

Figure 2.3: Impulse Responses to a Contractionary Monetary Policy Shock



Note: Horizontal axis shows quarters after a contractionary monetary policy shock of 10 basis points that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state for variables other than the interest rates and the loan margin, which are expressed in deviations from the steady state in percent points. The blue dashed line corresponds to perfect banking competition. The pink solid line corresponds to imperfect banking competition. Leverage ratio is computed as total asset value $(qk^E + q_h h^E)$ over net worth n and PED denotes the elasticity of the entrepreneur's loan demand to the loan rate.

Taylor rule at the beginning of period 1, so that the gross nominal deposit rate (policy rate) R_t is raised.

As can be seen from Figure 2.3, after a contractionary monetary policy shock, the responses of output, consumption, investment and physical capital tend to be amplified under imperfect banking competition. Following Andrés and Arce (2012), there are three relevant channels through which a monetary policy shock can affect this model economy: the traditional interest rate channel, the endogenous loan interest margin, and the net worth effect.

The interest rate channel works in a standard way via sticky prices. When a contractionary monetary shock raises the gross nominal interest rate R_t, the gross real interest rate also rises due to price rigidity. The increase in the real deposit rate reduces consumption via household intertemporal substitution. However, there is little difference in the response of the real deposit rate for the two types of banking competition, indicating that this channel is not very important in explaining the amplified responses of output and consumption under imperfect banking competition.

The amplification effect can be explained by the rise in the real loan interest margin. With perfect banking competition, the loan interest margin is zero, hence households and entrepreneurs face the same real interest rate. A higher real loan rate after the contractionary monetary policy shock directly increases the tightness of the binding collateral constraint (2.23). Besides, a higher real loan rate reduces the demand for physical capital and housing and thus their prices, leading to a lower value of collateralised assets and a further tightening of the binding constraint by reducing the entrepreneur's borrowing capacity. As a result, the Lagrange multiplier $\lambda^E_{2,t}$ rises by around 15% under perfect banking competition, indicating that the binding collateral constraint (2.23) tightens, as can be seen in Figure 2.3. This implies that the entrepreneur is more financially constrained and hence the loan demand becomes more inelastic. As can be seen from Figure 2.3, the elasticity of the market loan demand to the loan rate (PED) falls by more than 5%. With perfect banking competition, although the market loan demand becomes more inelastic, each individual bank faces a perfectly elastic loan demand and cannot affect the equilibrium loan rate. By contrast, with Cournot banking competition, each bank has market power and takes advantage of this reduction in loan demand elasticity by reducing the quantity of loans to achieve a higher equilibrium loan rate. Because of this joint effect between banks' market power and the time-varying market loan demand elasticity PED_t, the real loan rate rises further above that under perfect banking competition and the real loan margin increases by 5 basis points on impact, as can be seen from Figure 2.3. The higher real loan interest rate reduces the entrepreneur's demand for housing and physical capital and their prices by more, thus tightening the binding constraint even further. As shown in Figure 2.3, the Lagrange multiplier $\lambda^E_{2,t}$ rises by around 25% under imperfect banking competition.

The net worth effect tends to reinforce the amplification effect arising from the countercyclical real loan margin. With both types of banking competition, the fall in net worth after a contractionary monetary policy shock reduces entrepreneur's borrowing capacity and hence their demand for housing and physical capital. This depresses the real prices of housing and physical capital, and further reduces entrepreneur's net worth and their access to external financing, leading to a further fall in demand. This net worth effect is stronger under imperfect banking competition due to a larger fall in entrepreneur's net worth, as shown in Figure 2.3, which is likely due to a larger increase in the real loan rate and a larger fall in asset prices under imperfect banking competition.

2.4.2 Productivity Shock

After a one-standard-deviation negative productivity shock at the beginning of period 1, the responses of output, investment and physical capital are slightly amplified, which can be explained by the countercyclical loan interest margin and the net worth effect.

As discussed in Section 2.4.1, the rise in the real loan margin is due to the joint effect of banks' market power and the decreasing loan demand elasticity. Falling asset prices after the negative productivity shock reduce entrepreneurs' borrowing capacity and make them more financially constrained, thus resulting in more inelastic loan demand (PED falls), as can be seen from Figure 2.4. With imperfect banking competition, banks have market power and take advantage of the lower loan demand elasticity by reducing their loan quantities to achieve a higher equilibrium loan rate.

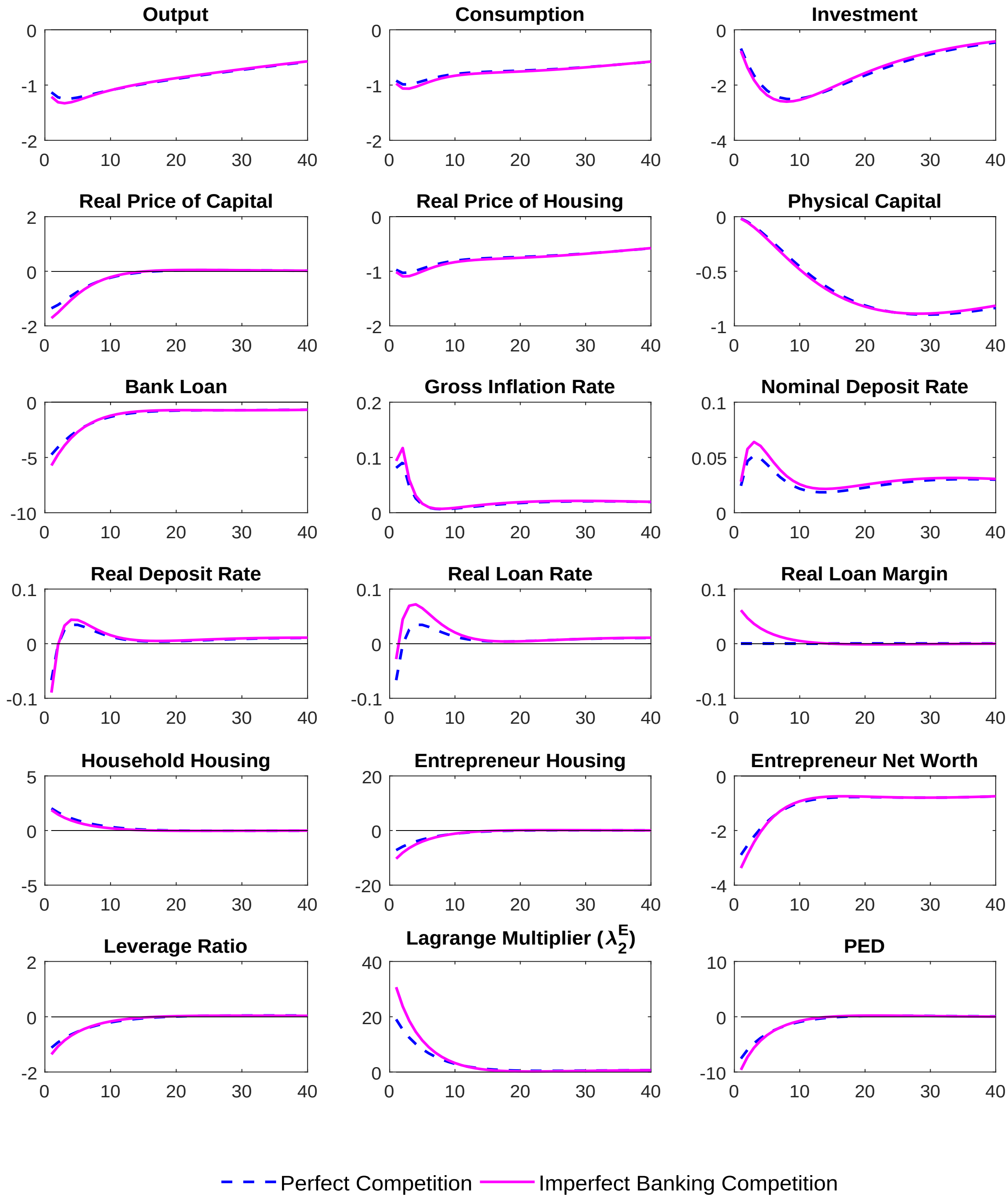
The net worth effect reinforces the amplification effect arising from the countercyclical loan margin. As discussed in Section 2.4.1, a lower net worth reduces the entrepreneur's borrowing capacity and hence their demand for physical capital and housing, which then depresses the asset prices and thereby further reduces the entrepreneur's net worth. With imperfect banking competition, the net worth effect is stronger as the fall in asset prices is larger (due to the countercyclical loan margin that reduces the entrepreneur's demand for housing and capital by more) and thereby the fall in the entrepreneur's net worth is greater, as shown in Figure 2.4.

Despite the countercyclical real loan margin and the net worth effect, the amplification effect of imperfect banking competition is weaker after the negative productivity shock. This is likely because of a debt-deflation effect. A negative productivity shock is a negative supply shock, so it is inflationary and the real loan rate falls due to the higher inflation, which is the main difference from the contractionary monetary policy shock.

Asset prices still decline after the negative productivity shock, which reduces the entrepreneur's borrowing capacity. But the real loan rate now falls, which tends to improve the entrepreneur's borrowing capacity. As a result, after the negative productivity shock, the tightness of the binding borrowing constraint increases to a lesser extent.[16]

[16] Note that one cannot compare the values of $\lambda_{2,t}^{E}$ between the productivity shock and the monetary policy shock. But notice that although both are one-standard-deviation shocks, the reduction in output

Figure 2.4: Impulse Responses to a Negative Productivity Shock



Note: Horizontal axis shows quarters after a one-standard-deviation negative productivity shock that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state for variables other than the interest rates and the loan margin, which are expressed in deviations from the steady state in percent points. The blue dashed line corresponds to perfect banking competition. The pink solid line corresponds to imperfect banking competition. Leverage ratio is computed as total asset value $(qk^E + q_h h^E)$ over net worth n and PED denotes the elasticity of the entrepreneur's loan demand to the loan rate.

2.4.3 Collateral or Macroprudential Policy Shocks

This section investigates the shocks to the pledgeability ratios, m_h and m_k, which can be interpreted as either collateral shocks or macroprudential policy shocks. These collateral shocks are supply-side shocks, as they directly affect the supply of credit to the entrepreneur and thereby the supply of output. Therefore, they tend to reduce output but increase inflation.

Negative Shock to m_h

After a one-standard-deviation ($\sigma_{m_h} = 0.01$) negative m_h shock at the beginning of period 1, the responses of output, investment and physical capital are all amplified under imperfect banking competition, which can be explained by the countercyclical real loan margin and the net worth effect. The mechanisms are similar to the negative productivity shock, as both are negative supply shocks. However, there are two main differences from the negative productivity shock.

First, the amplification effect is larger compared to that after the negative productivity shock because of the exogenous reduction in the pledgeability ratio m_h that directly reduces the fraction of the housing collateral that can be borrowed against and thereby lowers the entrepreneur's borrowing capacity through the binding collateral constraint (2.23). The decrease in $m_{h,t}$ makes the entrepreneur more financially constrained and thus reduces the loan demand elasticity. Although the negative shock to m_h is also inflationary, which tends to reduce the real debt burden and improve the entrepreneur's borrowing capacity, this debt-deflation effect is dominated by the effect of a lower pledgeability ratio m_h that directly reduces the borrowing capacity.

Second, the reduction in output after the negative shock to m_h is much smaller. This is because the pledgeability ratio affects the output only indirectly through the supply of credit, unlike the negative productivity shock that directly lowers the output.

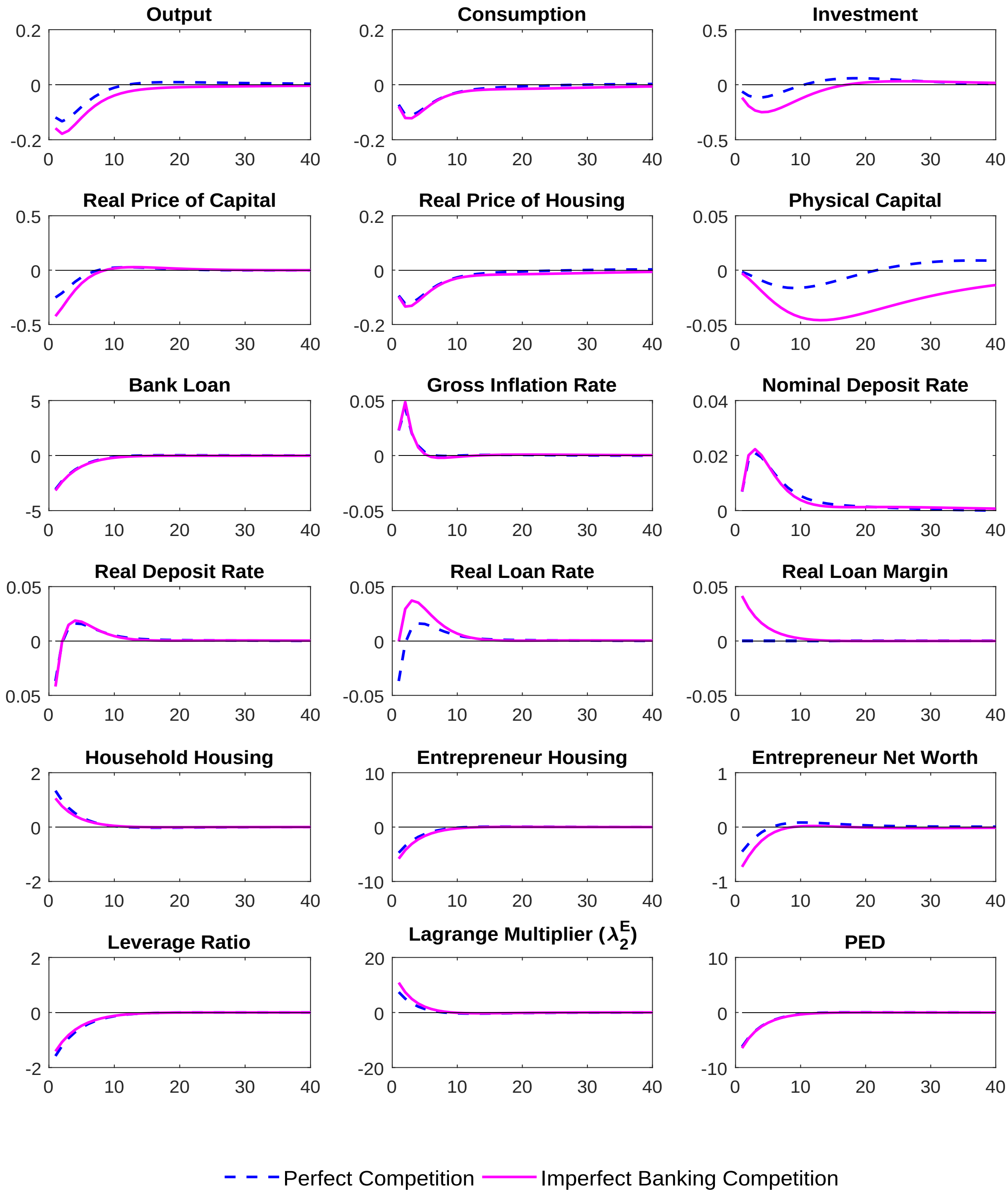
Negative Shock to m_k

In response to a one-standard-deviation ($\sigma_{m_k} = 0.01$) negative m_k shock at the beginning of period 1, imperfect banking competition is again found to amplify aggregate fluctuations, which can be explained by the countercyclical real loan margin and the net worth effect, as discussed in Section 2.4.2. The two differences from the negative productivity shock are the same as discussed for the negative shock to m_h. More specifically, the amplification effect after the negative shock to m_k is stronger since the lower pledgeability ratio directly reduces the entrepreneur's borrowing capacity and the magnitude of the output response is smaller since m_k only has an indirect effect on output through the supply of credit.

---

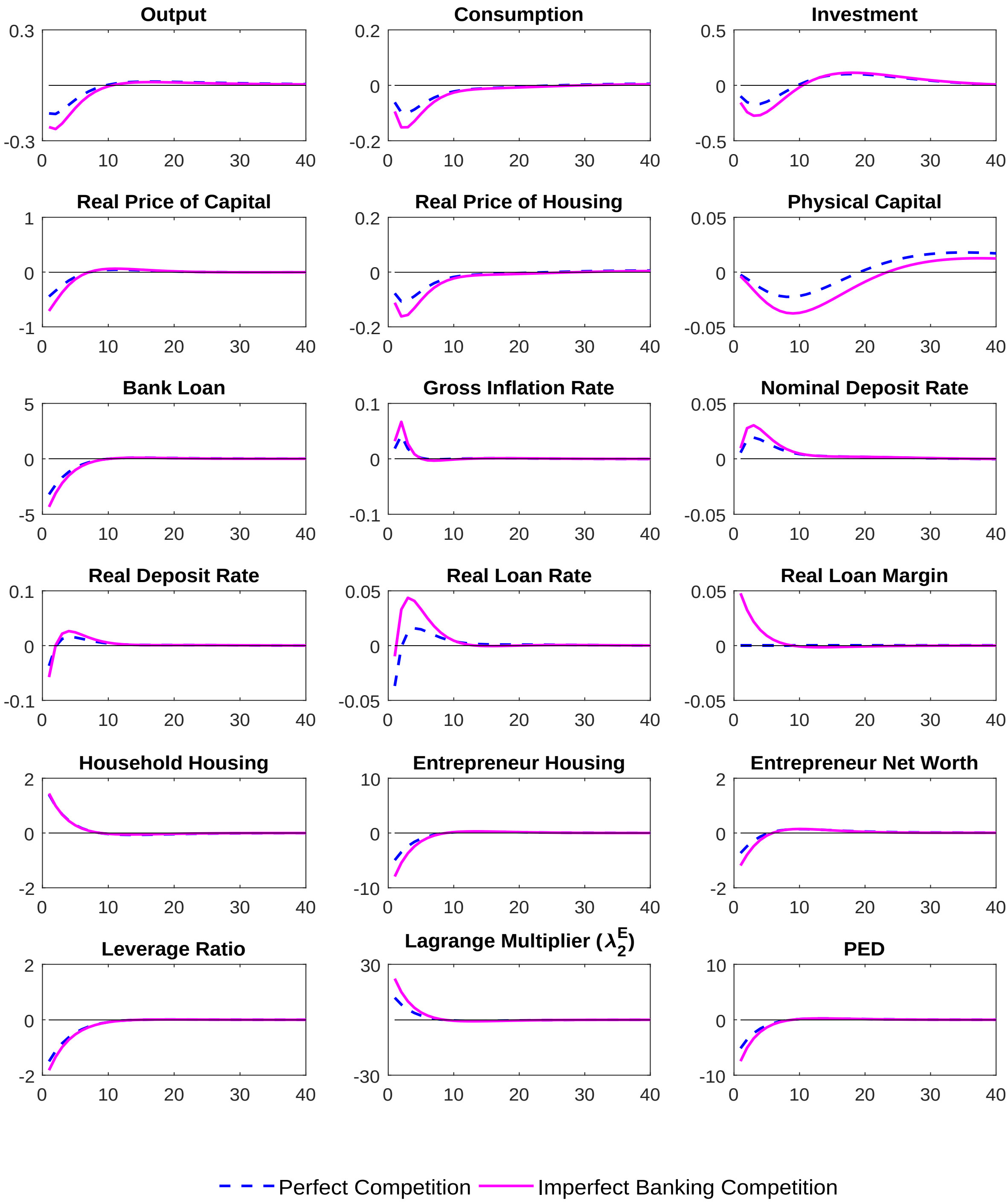
after the negative productivity shock is twice as large while the response for $\lambda^E_{2,t}$ is similar compared to those after the contractionary monetary policy shock .

Figure 2.5: Impulse Responses to a Negative m_h Shock



Note: Horizontal axis shows quarters after a one-standard-deviation negative shock to m_h that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state for variables other than the interest rates and the loan margin, which are expressed in deviations from the steady state in percent points. The blue dashed line corresponds to perfect banking competition. The pink solid line corresponds to imperfect banking competition. Leverage ratio is computed as total asset value $(qk^E + q_h h^E)$ over net worth n and PED denotes the elasticity of the entrepreneur's loan demand to the loan rate.

Figure 2.6: Impulse Responses to a Negative m_k Shock



Note: Horizontal axis shows quarters after a one-standard-deviation negative shock to m_k that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state for variables other than the interest rates and the loan margin, which are expressed in deviations from the steady state in percent points. The blue dashed line corresponds to perfect banking competition. The pink solid line corresponds to imperfect banking competition. Leverage ratio is computed as total asset value $(qk^E + q_h h^E)$ over net worth n and PED denotes the elasticity of the entrepreneur's loan demand to the loan rate.

However, the two types of collateral shocks are asymmetric and hence the impulse responses between the negative shocks to m_k and m_h are different.[17] As can be seen from Figure 2.5 and 2.6, the amplification effect after the negative shock to m_k is slightly stronger. This is because the standard deviations for both types of shocks are assumed to be the same. Since the steady state value of m_k is smaller than m_h, a one-standard-deviation shock leads to a larger proportional drop in m_k than m_h, which leads to a larger reduction in the entrepreneur's borrowing capacity. As a result, the entrepreneur becomes more financially constrained and thereby loan demand is more inelastic, leading to a slightly larger increase in the loan interest margin and thus the amplification effect.

2.5 Sensitivity Analysis

I check the robustness of the results by changing the number of banks N, the parameters in the Taylor rule (ρ_r, κ_π, and κ_y), the steady state pledgeability ratios (m_h and m_k), the investment adjustment cost parameter χ, the output elasticity of housing v and the household preference for housing ϕ_h, while each time, all the other variables are calibrated as in the baseline analysis. This section discusses the sensitivity of the baseline results in Section 2.4 to these changes of parameters in turn.

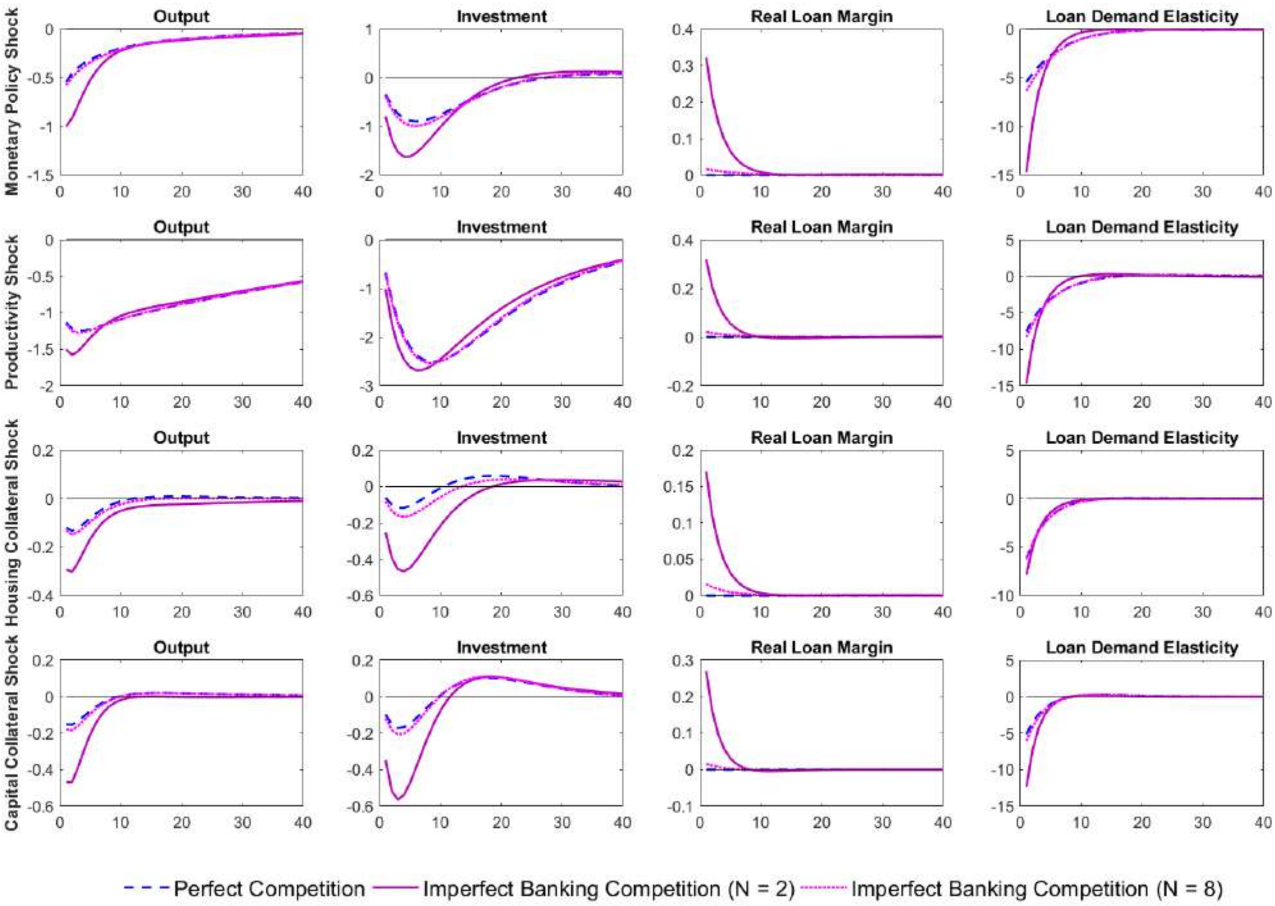
Figure 2.7 shows the impulse responses of output, investment, real loan margin and loan demand elasticity PED_t after four types of shocks when the number of banks N is two, eight and infinity (i.e., perfect competition). As can be seen, when there are only two banks, the amplification effect is much larger.[18] The real loan margin rises by around 30 basis points after the contractionary monetary policy shock, negative productivity shock and negative shock to m_k, which is almost six times as large compared to the baseline results with $N = 4$. Figure 2.7 shows that when N increases to eight, banks' market power is greatly reduced and the outcome is very similar to the perfect banking competition benchmark.

The baseline results are robust to eliminating interest rate smoothing by setting $\rho_r = 0$. In this case, the Taylor rule (2.44) reduces to the simplest possible form. As shown in Figure A.1 in Appendix A.6, the results are very similar to the baseline results in Section 2.4, apart from the monetary policy shock, after which the magnitude of the responses under both types of banking competition are much reduced. This is because the contractionary monetary policy no longer leads to a persistent increase in the nominal interest rate due to interest rate smoothing, thus reducing the effective size of the shock. As a result, the effects of the shock are substantially smaller.

[17] The two types of shocks are asymmetric due to differences in the steady state values of m_k and m_h, and the values of the capital share α and housing share v in the production function, as can be seen from the elasticities of the entrepreneur's housing demand PEH_t (A.56) and capital demand PEK_t (A.58) to the loan rate in Appendix A.3.2. In addition, since housing is inelastically supplied whereas capital can be adjusted, this gives rise to another source of asymmetry between the two types of shocks.

[18] When $N = 2$ (8), the steady state annualised net loan rate is around 10.6% (3.0%).

Figure 2.7: Impulse Responses for Different Shocks and Number of Banks N



Note: Horizontal axis shows quarters after the shock that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state for variables other than interest rates and the loan margin, which are expressed in deviations from the steady state in percent points. The blue dashed line corresponds to perfect banking competition. The pink solid line corresponds to imperfect banking competition. Each row shows the impulse responses of four variables after a given type of shock: a 10 basis-point contractionary monetary policy shock, a one-standard-deviation negative productivity shock, and one-standard-deviation negative shocks to m_h and m_k.

Increasing κ_π leads to greater response to the deviation of inflation from its target. Figure A.2 in Appendix A.6 shows that the fall in output is smaller under both types of banking competition after the contractionary monetary policy shock that is deflationary, but larger after the negative productivity shock that is inflationary. In the latter case, the real loan rate rises more to bring down inflation, causing a larger fall in output.

However, the amplification effect after a productivity shock in the baseline analysis is not robust to changes in the sensitivity κ_y of the policy rate to the output gap, for a given calibration of $\kappa_\pi = 1.5$. Figure A.3 in Appendix A.6 shows that when $\kappa_y = 0.2$, the loan margin actually decreases after a negative productivity shock. This is because after the negative productivity shock, the fall in output is very large. With $\kappa_y > 0$, the central bank's response to the output gap leads to a lower policy rate, and thereby a lower real loan rate. This makes the entrepreneur less financially constrained and the loan demand more elastic, causing a slight decline in the loan interest margin in this case.[19]

The results are robust to calibration for a range of different values for the pledgeability ratios. Keeping $m_h = 0.8$ as in the baseline calibration, the amplification effect is robust to varying m_k from 0.3 to 0.95, while keeping $m_k = 0.5$ as in the baseline calibration, the amplification effect is robust to varying m_k from around 0.5 to 0.9. In each case, a higher m_k or m_h tends to increase the amplification effect. This is because with a higher steady state value of m_h or m_k, the effect of changes in the real loan rate, capital, housing and asset prices on total borrowing is stronger because of the binding collateral constraint (2.23).[20] When m_k (m_h) is below 0.3 (0.5), the amplification is very weak. Figures A.4 and A.5 in Appendix A.6 show the impulse responses after four different types of shocks after setting m_k and m_h to their lower bounds respectively. As can be seen, in each case, the magnitude of the countercyclical real loan margin is much smaller. The responses of output after a contractionary monetary policy shock and a negative productivity shock are only slightly amplified. Intuitively, for a given rise in the loan rate or fall in asset prices, a higher m_k or m_h tends to amplify these changes, leading to a greater tightening of the collateral constraint and thus a larger fall in loan demand elasticity and a greater rise in the real loan margin.

Increasing the output elasticity of housing v from 0.05 to 0.15 has a larger effect on output, as shown in Figure A.6 in Appendix A.6. This is because when v is higher, any reduction in the entrepreneur's housing after a negative shock has a greater impact on output due to the production function (2.8). However, the amplification effect is smaller, because a higher v tends to increase the entrepreneur's housing-to-capital ratio, which means housing is used more intensively in the production, and hence the reduction in the entrepreneur's housing (one of the collateral assets) is smaller after the negative

[19]When $\kappa_y < 0.1$, the loan interest margin increases slightly.

[20]Note that changes in the steady state pledgeability ratio are different from collateral shocks. As discussed in Section 2.4.3, a negative collateral shock (that reduces the pledgeability ratio) can result in a lower loan demand elasticity and a countercyclical loan interest margin, which in turn amplifies aggregate fluctuations.

shocks. As a result, the entrepreneur's borrowing capacity is not reduced as much and the entrepreneur is less constrained compared to the case when v is 0.05, leading to a smaller fall in the elasticity of loan demand and thereby a smaller magnitude of the countercyclical real loan margin, as can be seen from Figure A.6.

Changing the parameter of the household's preference for housing ϕ_h from 0.1 to 0.3 changes the steady state allocation of housing between the household and the entrepreneur and does not affect the impulse responses much, as shown in Figure A.7 in Appendix A.6. This is because a stronger household's preference for housing leads to a higher household's demand for housing and thereby a higher real price of housing. However, the entrepreneur's housing demand is reduced as the total supply of housing is one. These two opposite effects tend to roughly offset each other, and hence changes in ϕ_h have little impact on the impulse responses.[21]

The amplification effect is not sensitive to changing the investment adjustment cost parameter χ. A higher χ mainly reduces the magnitude of the response for investment and physical capital. As shown in Figure A.8 in Appendix A.6, different values of χ give very similar responses of output, despite the noticeable differences in the responses of physical capital. This is because a one percentage reduction of capital from its steady state only leads to 0.33 (calibration for α is 0.33) percentage reduction of output from its steady state, given everything else the same.

Finally, one crucial assumption used in this paper is that borrowers are always financially constrained. Relaxing this assumption will lead to asymmetries between positive shocks and negative shocks. Starting from a steady state of a binding constraint, the results after negative shocks remain unchanged as the constraint will become more tightly binding. By contrast, large enough positive shocks can make the borrower financially unconstrained initially, leading to a potentially more elastic loan demand and procyclical loan margin during the initial periods.

2.6 Conclusions

This paper studies the effect of imperfect banking competition on macroeconomic volatility in a DSGE model embedded with the agency problem that gives rise to the collateral constraint. I find that in the presence of a binding collateral constraint, imperfect banking competition tends to amplify the responses of output, investment and physical capital after a contractionary monetary shock and negative collateral shocks, while this amplification effect is relatively weak after a negative productivity shock.

The amplification effect can be explained by the countercyclical loan interest margin, which arises from a joint effect between banks' market power and the time-varying loan demand elasticity facing the banks. After an adverse shock, a tightening of the binding

[21] Note that if the borrowing constraint were nonbinding and the entrepreneur rented housing, then $q_{h,t}h_t^E = \frac{vy_{w_t}}{x_t}$ and changes in $q_{h,t}$ and h_t^E would perfectly offset each other.

collateral constraint indicates that the borrowing capacity is reduced and the borrower is more financially constrained, thus leading to a more inelastic loan demand. Banks with market power can take advantage of this lower loan demand elasticity by reducing their loan quantities to achieve a higher equilibrium loan rate, thereby increasing the loan interest margin. This countercyclical real loan margin has a clear amplification effect on the response of output after a monetary policy shock and collateral shocks.

The results in this paper suggest that imperfect banking competition tends to be an important propagation mechanism for macroeconomic shocks especially when the degree of banking competition is low. This has implications for monetary policy and macroprudential policy. With imperfect banking competition and a binding collateral constraint, a contractionary monetary policy that makes borrowers more financially constrained can have greater impact on the real economy via a countercyclical loan margin. As a result, when policymakers decide by how much to raise the policy rate, it is important to factor in the amplification effect from imperfect banking competition.

In addition, the results also have implications for macroprudential policy in the form of adjustments in loan-to-value (LTV) ratios. The collateral shocks (i.e., shocks to the steady state pledgeability ratios on housing and physical capital) in this paper resemble macroprudential policy shocks. In the presence of a binding borrowing constraint, a temporary reduction in LTV ratio directly reduces the borrowing capacity of the entrepreneur and makes the entrepreneur more financially constrained, leading to a more inelastic loan demand and a countercyclical loan margin that amplifies aggregate fluctuations. Consequently, when changing the LTV ratio for financial stability concerns, the side effect on macroeconomic volatility should not be neglected.

Chapter 3

Imperfect Banking Competition and Financial Stability

3.1 Introduction

Does banking competition jeopardize financial stability? Understanding how banking competition affects financial stability provides crucial guidance on choosing the most effective macroprudential policy tools. Despite its importance, the relationship between banking competition and financial stability remains highly debated in the literature.

Much of the literature has focused on how bank competition affects banks' or borrowers' risk-taking.[1] Instead, this paper examines how competition affects banks' equity-to-assets ratios (equity ratios) and thereby financial stability measured through banks' default probabilities. By building a model of imperfect banking competition featuring the accumulation of bank equity via retained earnings, this paper finds that less banking competition can lead to a large gain in financial stability provided that banks retain the greater profits to build up their capital buffer.

Although less banking competition improves financial stability, it reduces aggregate output and hence macroeconomic efficiency, because a higher loan rate leads to a lower demand for physical capital and thus lower output. This paper quantifies the importance of the financial stability gain from less banking competition relative to the macroeconomic efficiency loss.

This paper shows that bank equity accumulation is important for understanding the trade-off between financial stability and macroeconomic efficiency. In the absence of bank equity accumulation, the financial stability gain from less banking competition is very small and is always outweighed by the macroeconomic efficiency loss. As a result, perfect banking competition is the best in this case. However, when banks retain their profits to build up their capital buffer over time, the financial stability gain from less banking competition can be large enough to outweigh the macroeconomic efficiency loss.

[1] See Corbae and Levine (2018), Boyd and De Nicolo (2005), Allen and Gale (2000), Keeley (1990), etc.

The importance of bank equity accumulation implies the relevance of macroprudential regulation on banks' dividend distribution.[2] For instance, by limiting banks' dividend distribution to shareholders, macroprudential policies can help to obtain a larger financial stability gain from less banking competition. Empirically, this paper finds supporting evidence for the model's prediction that when banks accumulate equity over time, less banking competition can lead to a large gain in financial stability measured by banks' default probabilities.

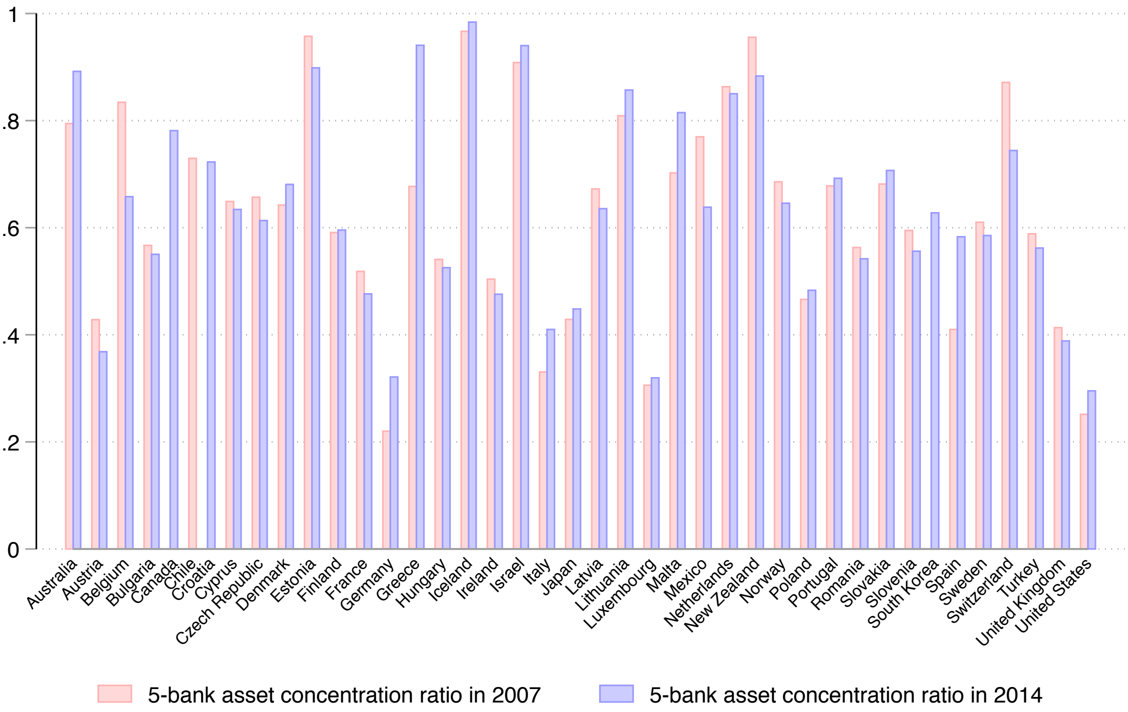
The imperfectly competitive nature of the banking sector can be seen in Figure 3.1, which shows that the largest 5 banks by total assets share more than 60% of the market in many EU and OECD countries in 2007 and 2014. This paper models imperfect banking competition via a Cournot banking sector where banks with different efficiencies compete for loans in each period. Loans are financed by deposits and equity accumulated via retained earnings. Entrepreneurs with limited liability and no initial wealth borrow via non-state-contingent debt contracts from the banking sector, to finance the purchase of physical capital for production. Entrepreneurs face idiosyncratic and aggregate shocks to productivity after installing the physical capital. Banks can perfectly diversify the idiosyncratic risk, but cannot diversify the aggregate risk, so they can default ex post if an adverse aggregate productivity shock causes too many entrepreneurs to default and their equity is not enough to absorb the loan losses. Hence, banks with higher equity ratios are better able to withstand aggregate shocks and have lower default probabilities. This paper analyzes how banking competition affects banks' equity ratios and thereby their default probabilities.

A key theoretical insight is that less banking competition can lead to a large gain in financial stability provided that banks retain the greater profits to build up their equity over time. With less banking competition, banks have higher profit margins, which provide a buffer against losses. However, this static margin effect only has a small impact on financial stability. When taking into account that banks can accumulate the greater profits over time, less banking competition can lead to a much larger gain in financial stability. This implies an important role for macroprudential regulation on banks' dividend distribution.

The model gives rise to some empirical implications that I assess using data for EU and OECD countries from 1999 to 2016. The model predicts that when banks retain their profits as equity over time, less banking competition improves financial stability measured through banks' default probabilities. I assess this prediction in two steps. First, based on the model, less banking competition leads to a larger change in bank equity when banks retain their profits. Second, banks with higher equity ratios have lower default probabilities. I provide two sets of supporting evidence. First, bank concentration, which is used as an inverse proxy for banking competition, has a significant positive effect on

[2]Macroprudential policies can regulate banks' dividend distribution. For example, in the US, banks that fail the stress test face restrictions on dividend distribution to shareholders.

Figure 3.1: Bank Concentration for EU and OECD Countries in 2007 and 2014



Note: The annual country-level 5-bank asset concentration ratio is the sum of market shares of the largest 5 banks by total assets. For EU countries, this is based on credit institutions (defined as receiving deposits or other repayable funds from the public and granting credits for its own account) authorized by a given country, using ECB data. For non-EU OECD countries (i.e., Australia, Canada, Chile, Iceland, Israel, Japan, Mexico, New Zealand, Norway, South Korea, Switzerland, Turkey and US), the 5-bank asset concentration ratio is computed using Bankscope annual data.
Data sources: ECB Macroprudential Database, Bankscope

the change in bank equity. Second, banks' equity ratios are negatively related to their default probabilities, proxied by credit default swap (CDS) spreads.[3] I also assess the model prediction in one step by looking at the direct relationship between banks' CDS spreads and bank concentration. I find that higher bank concentration leads to lower CDS spreads during the post-crisis period, which is consistent with the model prediction.[4]

The existing theoretical literature on the relationship between banking competition and financial stability can be classified into three categories: the competition-fragility view, competition-stability view, and an ambiguous relationship. The literature supporting the competition-fragility view tends to focus on the risk-taking channel (e.g., Corbae and Levine, 2018; Corbae and D'Erasmo, 2011; Allen and Gale, 2000; Hellmann et al., 2000; Matutes and Vives, 2000; Keeley, 1990) – competition reduces banks' franchise values (i.e., net present value of expected future profits) and thus induces more risk-taking by

[3]The CDS spread is the price of insurance against the default of a bank, so a higher CDS spread implies a higher bank default probability.

[4]When directly regressing banks' CDS spreads on bank concentration, the latter is only significant during the post-crisis period due to the lack of cross-country variation in CDS spreads during the pre-crisis period.

banks.[5] In contrast, there is also literature supporting the competition-stability view. For instance, by focusing on borrowers' risk-taking rather than banks' risk-taking, Boyd and De Nicolo (2005) introduce the risk-shifting hypothesis – competition lowers the loan rate and reduces borrowers' risk-taking, thus making banks' loan portfolio safer. Martinez-Miera and Repullo (2010) combine the risk-shifting effect with the margin effect that reduces profits and thereby the buffer against losses, and argue that the relationship between competition and stability is ambiguous, depending on which effect dominates.[6]

Similarly, the existing empirical evidence can also be classified according to three different views on the relationship between competition and stability: the competition-fragility view (e.g., Carlson et al., 2018; Corbae and Levine, 2018; Jiang et al., 2017; Beck et al., 2013; Ariss, 2010; Yeyati and Micco, 2007; Salas and Saurina, 2003; Keeley, 1990), the competition-stability view (e.g., Anginer et al., 2014; Dick and Lehnert, 2010; Uhde and Heimeshoff, 2009; Schaeck and Cihák, 2007), and an ambiguous relationship (e.g., Faia et al., 2018; Jiménez et al., 2013; Tabak et al., 2012). One reason to explain the mixed empirical results is that competition affects different types of risks differently, as pointed out by Freixas and Ma (2015).[7] In addition, the diversity of measures used for competition explains part of the mixed empirical results. In fact, there are papers that find that different measures for competition can lead to opposite results for the impact of competition on stability (e.g., Fu et al., 2014; Schaeck et al., 2009; Beck et al., 2006).

This paper contributes to the existing literature in three major respects. First, this paper introduces a new mechanism, the equity ratio effect, whereby competition affects banks' equity ratios and thereby their default probabilities. A few papers studying bank competition and financial stability also model bank equity but they look at the role of equity in deterring bank risk-taking (Corbae and Levine, 2018; Hellmann et al., 2000; Keeley, 1990), or making banks commit to monitoring (Allen et al., 2011). Instead, this paper focuses on the role of equity as a buffer against loan losses. More specifically, it incorporates the static margin effect and introduces dynamic bank equity accumulation via retained earnings. Based on the calibrated model, the static margin effect only has a

[5]Besides, bank competition can also jeopardize stability by worsening the coordination problem between depositors that can foster bank runs (Vives, 2016). Banks with long-term assets financed by short-term liabilities are vulnerable to runs, irrespective of competition, as in Diamond and Dybvig (1983). However, more intense competition raises the probability of failure in a symmetric interior equilibrium where banks are direct competitors for deposits (Matutes and Vives, 1996). Similarly, Egan et al. (2017) find that banks with high default probabilities are willing to offer high insured deposit rates. To compete for deposits, rival banks also raise their rates which reduce their margins and increase their default probabilities. This paper does not look at how competition affects stability via the bank-run channel by assuming full deposit insurance and a perfectly elastic supply of deposits.

[6]Caminal and Matutes (2002) also find that the relationship between bank competition and banking failures is ambiguous. In their model set-up, higher borrowers' investment implies a higher failure rate of the bank and under less banking competition, both the loan rate and the monitoring effort are higher, which affect borrowers' investment differently. While a higher loan rate reduces the investment, a higher monitoring effort raises the investment.

[7]For instance, Berger et al. (2009) find that banks with more market power have higher non-performing loans ratios, but have less overall risk measured by the Z-index, using more than 8000 banks in 23 developed countries from 1999 to 2005.

small impact on financial stability. However, when banks retain the greater profits to build up their capital buffer, less banking competition can lead to a much larger gain in financial stability. In essence, the improved profitability of banks under less banking competition is amplified over time.

Interestingly, if policymakers try to reduce competition to improve financial stability, the model suggests that it could make things worse in the short run. In particular, if a solvent bank merges with a distressed bank during a crisis, the merged bank with greater market power would have a larger size of loan assets, which reduces its equity ratio and hence raises its default probability. However, this short-run equity ratio effect tends to disappear over time due to faster equity accumulation with less banking competition, which results in higher bank equity ratios over time (long-run equity ratio effect).

Second, the paper provides a new measure to quantify the trade-off between financial stability and macroeconomic efficiency, using a calibrated version of the model.[8] More specifically, when there is little competition, the macroeconomic efficiency loss is very large. For instance, with a monopoly bank, aggregate output is 40% lower compared with a perfectly competitive banking sector. This large macroeconomic efficiency loss completely outweighs the financial stability gain. But when there are more than six banks, the financial stability gain from less banking competition becomes large enough over time to outweigh the macroeconomic efficiency loss, when banks engage in equity accumulation.

Since bank equity accumulation can result in a large gain in financial stability under imperfect banking competition, this implies an important role for macroprudential regulation that limits banks' dividend distribution to shareholders. Such macroprudential regulation of banks' dividend distribution has not received much attention in the literature, compared to capital requirements and deposit rate regulation,[9] even though it has already been implemented in practice, most notably by the US Federal Reserve for banks that fail stress tests. Admati et al. (2013) point out that prohibiting banks' dividend payouts for a period of time is an efficient and quicker way to have banks build up equity. The long-run equity ratio effect in this paper suggests a greater effectiveness of this macroprudential policy tool under less competition, because banks make higher profits.

Third, this paper provides new empirical evidence by assessing the model prediction that in the presence of bank equity accumulation, less banking competition improves financial stability. In this paper, I use bank concentration as an inverse measure for

[8]As Allen and Gale (2004) point out, although it is hard to measure the efficiency loss from concentration, it is unwise to neglect the efficiency costs. To address the balance between competition and stability, it is important to have a framework that allows for welfare analysis at different levels of competition. While Corbae and Levine (2018) compare the efficient level of risk taking and investment chosen by a social planner in a frictionless economy with the levels chosen in a decentralized Cournot equilibrium embedded with other frictions, I focus on the efficiency loss caused by imperfect banking competition in this paper.

[9]See Repullo (2004), Hellmann et al. (2000) and Besanko and Thakor (1992) for analysis on capital requirements and deposit rate ceilings.

competition based on the Cournot model and neglect that there may be a weak relationship between bank concentration and other banking competition measures empirically (e.g. Claessens and Laeven, 2004).[10] Based on the model, financial stability is measured by banks' default probabilities, so banks' CDS spreads are used to proxy for their default probabilities, with an additional benefit that this market-based measure is less likely to cause endogeneity problems, relative to the accounting-based measures, as noted by Anginer et al. (2014).[11] By assessing the model prediction in two steps, this paper provides supporting evidence for the mechanism behind the relationship between banking competition and financial stability via banks' equity ratios. Furthermore, this paper provides new evidence by investigating the direct relationship between banks' default probabilities and banking competition. I find that bank concentration, as an inverse proxy for banking competition, has a significantly negative effect on banks' CDS spreads during the post-crisis period.

The remainder of the paper is structured as follows. Section 3.2 presents the model set-up and the basic model results. Section 3.3 explains the model calibration. Section 3.4 uses the calibrated model to illustrate the long-run and short-run equity ratio effects, and to quantify the relative importance of the macroeconomic efficiency loss and the financial stability gain associated with imperfect banking competition. Section 3.5 documents the data sources used in this paper. Section 3.6 discusses the empirical specifications and reduced-form results supporting the model predictions. Section 3.7 concludes.

3.2 Model

A model of non-state-contingent debt contracts between competitive entrepreneurs and a Cournot banking sector is presented in this section. Entrepreneurs are born each period and only live for two periods. They start with no initial wealth and hence need to borrow from banks at a fixed loan rate to purchase and install physical capital in their first period, which is used as the only input for production in their second period, at the end of which they consume the profits. Entrepreneurs with limited liability are assumed to be identical ex ante but their productivity is subject to an idiosyncratic shock and an aggregate shock in their second period. Banks with different efficiencies compete in loan

[10] The competition measures such as HHI (Herfindahl-Hirschman Index) and 5-bank concentration ratios are based on the Structure-Conduct-Performance (SCP) hypothesis (Bikker et al., 2012), which holds under Cournot competition. Under the SCP hypothesis, a highly concentrated banking industry tends to cause banks to behave in a non-competitive way to make higher profits, for instance, via non-competitive pricing. This structural measure (e.g., bank concentration measures) has been questioned a lot in terms of how well they capture competition (Bolt and Humphrey, 2015), which has led to the development of non-structural measures such as the Lerner index, Panzar-Ross H-statistic and Boone indicator, etc. However, each of these measures has also received criticism, as can be seen from Carbó et al. (2009), Bikker et al. (2012) and Schiersch and Schmidt-Ehmcke (2010), etc.

[11] Few papers in this literature use CDS data. A recent paper by Faia et al. (2018) only use CDS data on 15 global systematically important banks, while this paper covers 157 banks in EU and OECD countries.

quantities *à la* Cournot and all banks' loan quantities then determine the equilibrium loan rate. Entrepreneurs that suffer adverse productivity shocks may not be able to repay their loans, in which case banks would incur a collection cost or auditing cost to observe and verify their realized output, and then confiscate the output. Due to the large number of entrepreneurs that each receives a different idiosyncratic shock to productivity, banks can perfectly diversify the idiosyncratic loan risk. However, banks are all affected by the aggregate shock to productivity. Ex post, some banks may default after an adverse aggregate shock if their efficiency level or equity ratio is sufficiently low.

3.2.1 Entrepreneur's Problem

There is a unit mass of ex ante identical entrepreneurs indexed by $i \in [0,1]$ with no initial wealth. Each borrows from a bank at a non-state-contingent gross loan rate $R_{b,t}$ to purchase physical capital $k_{i,t}$ in period t. Entrepreneurs take the loan rate $R_{b,t}$ set by the banking sector as given. There is a common deterministic productivity level $A > 0$ that is subject to multiplicative shocks that only realize at the beginning of period $t+1$ after entrepreneurs have installed the capital. The idiosyncratic multiplicative shock $\omega \geqslant 0$ is i.i.d. across entrepreneurs and time, with a continuous c.d.f. $F(\omega)$ and $\mathrm{E}(\omega) = 1$. The aggregate multiplicative shock $\epsilon \geqslant 0$ has a continuous c.d.f. $\Gamma(\epsilon)$ and $\mathrm{E}(\epsilon) = 1$. Ex post, each entrepreneur i receives a different realized idiosyncratic shock $\omega_{i,t+1}$ and produces output $y_{i,t+1}$:

$$y_{i,t+1} = \omega_{i,t+1}\epsilon_{t+1}Ak_{i,t}^{\alpha} \tag{3.1}$$

where $\alpha \in (0,1)$ is the output elasticity of capital. Facing the same $R_{b,t}$, each entrepreneur has the same demand for physical capital, so $k_{i,t} = k_t \ \forall i$.

If the realized idiosyncratic productivity shock at the beginning of period $t+1$ is below a certain threshold $\bar{\omega}_{t+1}$, the entrepreneur is not able to repay the debt obligation $R_{b,t}k_t$, where $\bar{\omega}_{t+1}$ is determined by the following break-even condition:

$$\bar{\omega}_{t+1}\epsilon_{t+1}Ak_t^{\alpha} - R_{b,t}k_t \equiv 0 \quad \rightarrow \quad \bar{\omega}_{t+1} \equiv \frac{R_{b,t}k_t^{1-\alpha}}{\epsilon_{t+1}A} \tag{3.2}$$

As can be seen from (3.2), a higher $R_{b,t}$ leads to a higher entrepreneur's default threshold $\bar{\omega}_{t+1}$, keeping everything else unchanged, i.e., $\frac{\partial \bar{\omega}_{t+1}}{\partial R_{b,t}} > 0$. A higher realized aggregate productivity shock ϵ_{t+1} results in a lower default threshold $\bar{\omega}_{t+1}$, meaning that the proportion of defaulting entrepreneurs is smaller, i.e., $\frac{\partial \bar{\omega}_{t+1}}{\partial \epsilon_{t+1}} < 0$.

Both the idiosyncratic and aggregate shocks are unobserved ex ante (when entrepreneurs and banks are making their decisions). Ex post, entrepreneurs and banks can observe the realized aggregate shock ϵ_{t+1}. Each entrepreneur i can also observe the realized idiosyncratic shock $\omega_{i,t+1}$ ex post, but other agents need to incur an auditing cost or collection cost to observe it. Given the information asymmetry and a positive auditing cost, the optimal debt contract takes the form of a standard non-state-contingent debt contract (Gale and Hellwig, 1985). That is, the entrepreneur pays $R_{b,t}k_t$ when the repayment can be afforded (i.e., when $\omega_{i,t+1} \geqslant \bar{\omega}_{t+1}$). If the realized output is too low to cover the debt repayment (i.e., when $\omega_{i,t+1} < \bar{\omega}_{t+1}$), the entrepreneur declares bankrupt. Since each entrepreneur borrows from only one bank, the bank then verifies the defaulting entrepreneur's output, incurring a collection cost $\mu \in (0,1)$ that is proportional to the realized output, and seizes the output.

A larger capital stock k_t requires higher productivity to break even due to the diminishing marginal product of capital, so it leads to a higher default threshold $\bar{\omega}_{t+1}$ and thus raises the entrepreneur's default probability $F(\bar{\omega}_{t+1})$, keeping everything else the same:

$$\frac{\partial \bar{\omega}_{t+1}}{\partial k_t} = \frac{(1-\alpha)R_{b,t}k_t^{-\alpha}}{\epsilon_{t+1}A} > 0 \tag{3.3}$$

The representative entrepreneur takes the gross loan rate $R_{b,t}$ as given and chooses k_t to maximize expected profits, taking into consideration the effect of k_t on the default threshold $\bar{\omega}_{t+1}$. Hence, the entrepreneur with limited liability maximizes the following expected profit with respect to k_t:

$$\mathrm{E}_t\left[\int_{\bar{\omega}_{t+1}(R_{b,t},k_t,\epsilon_{t+1})}^{\infty} \omega\epsilon_{t+1}Ak_t^{\alpha}dF(\omega) - \int_{\bar{\omega}_{t+1}(R_{b,t},k_t,\epsilon_{t+1})}^{\infty} R_{b,t}k_t dF(\omega)\right] \tag{3.4}$$

where the expectation operator $\mathrm{E}_t[.]$ is taken over the distribution of the aggregate shock ϵ_{t+1}, and the entrepreneur's default threshold $\bar{\omega}_{t+1}$ is a function of the gross loan rate $R_{b,t}$, physical capital k_t and the aggregate shock ϵ_{t+1} (as explained above). The optimal loan demand k_t decreases with $R_{b,t}$, as shown in Appendix B.1.1, so the loan demand curve is downward-sloping:

$$\frac{dk_t}{dR_{b,t}} = -\frac{k_t}{(1-\alpha)R_{b,t}} < 0 \tag{3.5}$$

The banking sector affects the demand for loans via the equilibrium loan rate. In addition, the loan rate may also affect the entrepreneur's default threshold. However, in this model setup, $\bar{\omega}_{t+1}$ is independent of the gross loan rate $R_{b,t}$, as proved in Appendix B.1.2:

$$\frac{d\bar{\omega}_{t+1}}{dR_{b,t}} = \frac{\partial \bar{\omega}_{t+1}}{\partial R_{b,t}} + \frac{\partial \bar{\omega}_{t+1}}{\partial k_t}\frac{dk_t}{dR_{b,t}} = 0 \tag{3.6}$$

The positive partial effect of $R_{b,t}$ on $\bar{\omega}_{t+1}$ is analogous to the argument made by Boyd and De Nicolo (2005) that an increase in loan rate caused by less loan market competition can reduce borrowers' profitability, inducing them to choose a higher riskiness attached to their portfolio, which undermines financial stability.[12] The main difference is that by

[12]When riskiness itself is a choice variable, as commonly seen in the literature (Martinez-Miera and Repullo, 2010; Boyd and De Nicolo, 2005), and when the expected revenue from the debt-financed project is strictly increasing in the riskiness, the only way to make profit facing a higher loan rate is to choose a higher riskiness.

separating the choice variable k_t from the riskiness measure $\bar{\omega}_{t+1}$, this model gives rise to the possibility that the adverse impact of the loan rate on borrowers' profitability is internalized by the borrowers themselves such that there is no overall impact of $R_{b,t}$ on $\bar{\omega}_{t+1}$. In essence, this is because entrepreneurs facing a higher interest rate would reduce their loan demand. As shown in (3.6), the positive partial effect of $R_{b,t}$ on $\bar{\omega}_{t+1}$ is exactly offset by the effect of the reduction in loan demand in response to a higher loan rate, so banks do not affect the entrepreneur's default threshold. The result that $\frac{d\bar{\omega}_{t+1}}{dR_{b,t}} = 0$ holds more generally if the entrepreneur has full liability.[13] The fact that the entrepreneur's default probability is unaffected by the loan rate in this model greatly simplifies the problem of the Cournot banking sector.

3.2.2 Cournot Banking Sector

There are N risk-neutral banks with different marginal costs competing in loan quantities *à la* Cournot. Banks are indexed by j, where $j = 1, 2, 3, ..., N$. When $N = 1$, the banking sector consists of a monopoly bank and when N approaches infinity, the banking sector is perfectly competitive. In equilibrium, the total loan demand k_t from the entrepreneur's problem is equal to the total loan supply which is provided by j banks, such that $k_t = \sum_j k_{j,t}$, where $k_{j,t}$ denotes the loan quantity supplied by bank j in period t.

Banks diversify to reduce idiosyncratic risk by lending to a fraction $\frac{k_{j,t}}{k_t}$ of randomly selected ex ante identical entrepreneurs. Once the idiosyncratic and aggregate shocks realize, entrepreneurs with realized values of $\omega_{i,t+1}(\epsilon_{t+1})$ above the threshold $\bar{\omega}_{t+1}(\epsilon_{t+1})$ would be able to repay the full debt obligation and each bank j gets the loan repayment $\int_{\bar{\omega}_{t+1}(\epsilon_{t+1})}^{\infty} R_{b,t} k_{j,t} dF(\omega)$ from these non-defaulting entrepreneurs. The entrepreneurs with realized values $\omega_{i,t+1}(\epsilon_{t+1})$ below the threshold $\bar{\omega}_{t+1}(\epsilon_{t+1})$ will declare bankruptcy. In this case, banks verify and confiscate the output after incurring a collection cost, which is a fraction μ of the realized output. So bank j obtains $\frac{k_{j,t}}{k_t}(1-\mu)\int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})} \epsilon_{t+1}\omega A k_t^{\alpha} dF(\omega)$ from the defaulting entrepreneurs in its portfolio. Due to the large number of entrepreneurs, banks can perfectly diversify the idiosyncratic risk. The aggregate shock ϵ_{t+1} to productivity that hits all entrepreneurs, however, affects the fraction of entrepreneurs that default and thereby the fraction of nonperforming loans and banks' default probabilities. Consequently, banks with low efficiencies or low equity ratios can default ex post due to an adverse aggregate shock.

Assume bank j finances its loans $k_{j,t}$, which are the only assets on its balance sheet, via deposits and net worth (equity) $n_{j,t}$, which is accumulated through retained earnings. Assume there is a perfectly elastic supply of deposits at the exogenous gross deposit rate $R_t > 0$. Depositors are protected by a deposit guarantee from the government, who repays any depositors affected by bank default. Based on the balance sheet identity that

[13] It is shown in Appendix B.1.2 that with full liability of the entrepreneur, the default threshold at the optimal k_t is $\bar{\omega}_{t+1} = \frac{1}{\epsilon_{t+1}}$, which is also independent of $R_{b,t}$. Compared with full liability, the entrepreneur chooses a larger k_t under limited liability, leading to a higher default probability $F(\bar{\omega}_{t+1})$.

assets equal the sum of liabilities (deposits) and equity, the amount of deposits taken by bank j is $(k_{j,t} - n_{j,t})$. Each bank j has a different time-invariant marginal intermediation cost for loans $\tau_j \in (0,1)$, with higher τ_j indicating lower efficiency. Consequently, banks have different market shares in the Cournot equilibrium, with more efficient banks gaining higher market shares.

Let $\pi^B_{j,t+1}$ denote the net profit earned by bank j on period-t loans in period $t+1$. Assume bankers are appointed for one loan cycle, so they only care about maximizing the expected profit $\mathrm{E}_t\pi^B_{j,t+1}$ by choosing the loan quantity $k_{j,t}$.[14] Although bankers are short-lived, banks are long-lived and they can accumulate equity over time. The net profit of bank j in period $t+1$ depends on the aggregate shock ϵ_{t+1}:

$$\begin{aligned}\pi^B_{j,t+1} = &\int_{\bar{\omega}_{t+1}(\epsilon_{t+1})}^{\infty} R_{b,t}k_{j,t}dF(\omega) + \frac{k_{j,t}}{k_t}(1-\mu)\int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})} \epsilon_{t+1}\omega A k_t^{\alpha} dF(\omega) \\ &- R_t(k_{j,t} - n_{j,t}) - \tau_j k_{j,t} - n_{j,t}\end{aligned} \tag{3.7}$$

where the first RHS term represents the revenue from performing loans and the second term equals the revenue from nonperforming loans, both for a given level of the aggregate shock. The third RHS term is the gross deposit interest payment, and $\tau_j k_{j,t}$ equals bank j's intermediation cost. The gross loan rate $R_{b,t}$ is a function of bank j's loan quantity and all the other banks' loan quantities. Under Cournot competition, each bank j chooses its loan quantity $k_{j,t}$ to maximize its expected net profit, taking into account the impact of its loan quantity choice on $R_{b,t}$ and taking all the other banks' loan quantities as given. The equilibrium loan rate is determined by all banks' loan quantities.

Using the expression for $\bar{\omega}_{t+1}$ (3.2), it is shown in Appendix B.2.1 that the net profit (3.7) can be simplified to:

$$\pi^B_{j,t+1} = G(\epsilon_{t+1})R_{b,t}k_{j,t} - R_t(k_{j,t} - n_{j,t}) - \tau_j k_{j,t} - n_{j,t} \tag{3.8}$$

where $G(\epsilon_{t+1}) \equiv [1 - F(\bar{\omega}_{t+1}(\epsilon_{t+1}))] + \frac{1-\mu}{\bar{\omega}_{t+1}(\epsilon_{t+1})}\int_0^{\bar{\omega}_{t+1}(\epsilon)} \omega f(\omega)d\omega < 1$ can be interpreted as the fraction of the contractual gross loan revenue $R_{b,t}k_{j,t}$ that can be obtained by bank j for a given level of ϵ_{t+1}. The revenue fraction $G(\epsilon_{t+1})$ is smaller than one due to the nonperforming loans. The net profit of bank j can be negative if the realization of the aggregate shock in period $t+1$ is sufficiently low, more precisely, below a threshold $\bar{\epsilon}_{j,t+1}$. Although the aggregate shock is common to all banks, the default threshold $\bar{\epsilon}_{j,t+1}$ differs across banks due to different levels of equity $n_{j,t}$ and efficiency indicated by τ_j.[15] A higher bank's default threshold $\bar{\epsilon}_{j,t+1}$ implies a higher default probability for the bank.

[14] Since bankers are appointed for one loan cycle, they do not consider the effect of the loan quantity choice $k_{j,t}$ on the bank's survival probability in future periods. In a dynamic Cournot model where bankers take into account the effect of $k_{j,t}$ on the bank's default probability, each bank would choose a smaller $k_{j,t}$ to reduce its default probability, resulting in a higher equilibrium loan rate and a higher profit. So under dynamic Cournot, banks can accumulate equity faster due to higher profits.

[15] In the model, banks have no debt, but the "default threshold" of a bank refers to the threshold at which it goes bankrupt and "defaults" on its liabilities (deposits).

Bank Equity Accumulation

Equity in period $t+1$, $n_{j,t+1}$, is modelled as the retained earnings of the continuing bank j, which is the sum of $n_{j,t}$ and net profit $\pi^B_{j,t+1}$ net of any dividend payments $D_{j,t+1}$:

$$n_{j,t+1} = n_{j,t} + \pi^B_{j,t+1} - D_{j,t+1} \tag{3.9}$$

where $\pi^B_{j,t+1}$ is given by (3.8). As can be seen from (3.9), macroprudential regulation on banks' dividend distribution can affect the equity accumulation via $D_{j,t+1}$, leading to different dynamics of equity over time and thus affecting banks' equity ratios under a given level of competition. This section shows three different bank dividend distribution or macroprudential policies: (i) no dividend distribution; (ii) distribute all positive net profits; (iii) distribute only if the equity ratio exceeds a desired or required level. The effects of these three policies on equity accumulation are shown below.

Bank j's default threshold $\bar{\epsilon}_{j,t+1}$ is determined by the condition that the pre-dividend net worth (equity) in period $t+1$ is zero, i.e., $\pi^B_{j,t+1} + n_{j,t} = 0$. If the loss made by bank j ($\pi^B_{j,t+1} < 0$) is too large to be absorbed by its capital buffer $n_{j,t}$, then bank j goes bankrupt. Hence, a larger net worth $n_{j,t}$ lowers bank j's default threshold $\bar{\epsilon}_{j,t+1}$. The negative relationship between banks' equity ratios and their default thresholds is established in Section 3.2.3.

Case I: No dividend distribution

Assuming banks do not distribute to shareholders (i.e., $D_{j,t+1} = 0$), equity accumulates as follows:

$$n_{j,t+1} = n_{j,t} + \pi^B_{j,t+1} \tag{3.10}$$

which is the sum of the equity in the previous period and the realized net profit. Conditional on a non-negative $n_{j,t+1}$ at the beginning of $t+1$, the continuing bank j will then choose loan quantity $k_{j,t+1}$ to maximize $\mathrm{E}_{t+1}\pi^B_{j,t+2}$.

Case II: Distribute all positive net profits to shareholders

Assume that whenever bank j makes a positive net profit ex post, it will distribute all the net profit to its shareholders, so the dividend payment in period $t+1$ before choosing the loan quantity $k_{j,t+1}$ is:

$$D_{j,t+1} = \max\{\pi^B_{j,t+1},\ 0\} \tag{3.11}$$

where $\pi^B_{j,t+1}$ is the net profit of bank j for a given realized aggregate shock ϵ_{t+1}. According to the evolution of equity (3.9), the post-dividend equity of bank j in period $t+1$ is then:

$$n_{j,t+1} = \min\{n_{j,t} + \pi^B_{j,t+1},\ n_{j,t}\} \tag{3.12}$$

When the realized net profit $\pi^B_{j,t+1}$ is negative, equity capital $n_{j,t}$ is used to absorb this loss, and no dividend is paid to shareholders. As long as $n_{j,t+1}$ is non-negative, bank j can stay in the market and choose the loan quantity $k_{j,t+1}$, financed by the post-dividend equity $n_{j,t+1}$ (3.12) and deposits.

Case III: Distribute if equity ratio exceeds the desired or required level

Assume banks have a desired or required equity ratio κ^* and they only pay dividend when the pre-dividend equity $n_{j,t} + \pi^B_{j,t+1}$ exceeds the desired/required level $\kappa^* k_{j,t}$.[16] When the pre-dividend equity ratio $\frac{n_{j,t}+\pi^B_{j,t+1}}{k_{j,t}}$ falls short of κ^*, banks do not pay any dividend in period $t+1$ and instead, they keep accumulating their equity. Hence, the dividend payment made by bank j in period $t+1$ is:

$$D_{j,t+1} = \max\{n_{j,t} + \pi^B_{j,t+1} - \kappa^* k_{j,t},\ 0\} \tag{3.13}$$

According to the evolution of equity capital (3.9), bank j's equity in period $t+1$ after paying dividend (3.13) is then:

$$n_{j,t+1} = \min\{n_{j,t} + \pi^B_{j,t+1},\ \kappa^* k_{j,t}\} \tag{3.14}$$

Compared to Case II, even when the net profit $\pi^B_{j,t+1}$ is positive, if the pre-dividend equity $n_{j,t} + \pi^B_{j,t+1}$ is lower than the desired or required level as indicated by $\kappa^* k_{j,t}$, no dividend will be paid to the shareholders.

3.2.3 Basic Model Results

This section presents the basic model results and uses these results to show the macroeconomic efficiency loss from imperfect banking competition, the equity ratio effect, and the negative relationship between banks' equity ratios and their default thresholds.

Macroeconomic Efficiency Loss from Imperfect Banking Competition

Before any shocks realize, N heterogeneous banks with different levels of efficiency indicated by τ_j compete in loan quantities and the equilibrium loan rate is determined by all banks' choices of loan quantities. It is shown in Appendix B.2.1 that the equilibrium loan rate can be found by first taking the first-order condition of (3.7) with respect to $k_{j,t}$ for each bank j and then summing over all N banks' first order conditions. The equilibrium gross loan rate $R^*_{b,t}$ is:

$$R^*_{b,t} = \frac{R_t + \bar{\tau}}{\left(1 - \frac{1-\alpha}{N}\right) \mathrm{E}_t[G(\epsilon_{t+1})]} \tag{3.15}$$

[16] One example of this desired equity ratio is the capital ratio set by regulatory authorities.

where $G(\epsilon_{t+1}) \equiv \Big[[1 - F(\bar{\omega}_{t+1}(\epsilon_{t+1}))] + \frac{1-\mu}{\bar{\omega}_{t+1}(\epsilon_{t+1})} \int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})} \omega f(\omega) d\omega\Big] < 1$ denotes the fraction of $R_{b,t}k_{j,t}$ that can be obtained by bank j for a given level of aggregate shock ϵ_{t+1}, as can be seen in (3.8). This fraction is smaller than one due to the presence of defaulting entrepreneurs. A higher $\mathrm{E}_t[G(\epsilon_{t+1})]$ implies a smaller proportion of entrepreneurs are expected to default, which lowers $R^*_{b,t}$ due to less risk compensation. The parameter $\bar{\tau} \equiv \frac{1}{N}\sum_{j=1}^{N} \tau_j$ denotes the mean marginal intermediation cost across all banks. A higher $\bar{\tau}$ implies lower bank efficiency and raises $R^*_{b,t}$ due to a higher marginal cost. It can be seen from (3.15) that the equilibrium loan rate is larger than $R_t + \bar{\tau}$ due to the market power of banks for finite N and the presence of non-performing loans such that $\mathrm{E}_t[G(\epsilon_{t+1})]$ is smaller than one. With perfect banking competition (i.e., when N approaches infinity), the equilibrium loan rate $R^{PC}_{b,t}$ is:

$$R^{PC}_{b,t} = \frac{R_t + \bar{\tau}}{\mathrm{E}_t[G(\epsilon_{t+1})]} \tag{3.16}$$

which is lower than $R^*_{b,t}$, but still larger than the marginal cost $R_t + \bar{\tau}$ due to the presence of non-performing loans.

The marginal intermediation costs for the N banks are randomly drawn from a given distribution which is assumed to be time-invariant.[17] If τ_j is too high relative to the distribution mean $\bar{\tau}$, then bank j is too inefficient to operate profitably. It is shown in Appendix B.2.2 that the following condition on τ_j is sufficient to ensure that all banks are able to make a positive expected net profit for $R_t \geqslant 1$:

$$R_t + \tau_j < \frac{R_t + \bar{\tau}}{\left(1 - \frac{1-\alpha}{N}\right)} \tag{3.17}$$

Note that this condition is satisfied if all banks are identical so that $\tau_j = \bar{\tau}\ \forall j$. Assume that τ_j is randomly drawn from a given time-invariant distribution for all levels of N, so changes in N do not affect the distribution mean $\bar{\tau}$.[18] All results in the rest of this section are proved under this assumption and condition (3.17). Some fundamental properties of the Cournot equilibrium are summarised in the following proposition which is proved in Appendix B.2.3.

Proposition 1: *A higher number of banks N*
(i) reduces the equilibrium loan rate $R^*_{b,t}$: $\frac{dR^*_{b,t}}{dN} = -\frac{(1-\alpha)R^*_{b,t}}{N(N-1+\alpha)} < 0$;
(ii) increases the equilibrium aggregate loan quantity k^*_t: $\frac{dk^*_t}{dN} = \frac{k^*_t}{N(N-1+\alpha)} > 0$;
(iii) improves macroeconomic efficiency measured through higher expected output $A(k^*_t)^\alpha$.

[17] In simulation results shown in Section 3.4, τ_j is drawn from a reverse bounded Pareto distribution in order to produce an unequal distribution for equilibrium market shares with a few large banks and a lot of small banks.

[18] In essence, this assumes constant returns to scale for banks.

As the number of banks N increases (more intense banking competition), the equilibrium loan rate is lower, which raises the demand for physical capital and thus leads to a higher equilibrium aggregate loan quantity. Let k_t^{PC} denote the aggregate physical capital or loan quantity under perfect banking competition when the loan rate is $R_{b,t}^{PC}$ (3.16). Proposition 1 shows that the expected output under perfect banking competition $\mathrm{E}_t(y_{t+1}^{PC}) = A\left(k_t^{PC}\right)^{\alpha}$ is higher than that under imperfect banking competition due to a lower loan rate and hence a higher demand for physical capital. In addition, less banking competition leads to a larger macroeconomic efficiency loss (or loss in expected output) compared to perfect banking competition.

Equity Ratio Effect

This paper introduces a new mechanism, the equity ratio effect, which describes how competition affects banks' equity ratios $\frac{n_{j,t}}{k_{j,t}}$ and thereby banks' default probabilities. This mechanism differentiates between the short-run and long-run effects of banking competition on financial stability. The short-run equity ratio effect is a denominator effect via which banking competition changes the size of loan assets $k_{j,t}$, whereas the long-run equity ratio effect is a numerator effect via which banking competition affects the speed of equity accumulation and thereby the level of $n_{j,t}$ over time. This section explains the short-run and long-run equity ratio effects in turn. To analyze the former effect, it is necessary to show how bank j's loan quantity $k_{j,t}$ changes with N, while for the latter, it is important how bank j's net profit $\pi_{j,t}^{B}$ changes with N.

It is shown in Appendix B.2.4 that each bank j's optimal equilibrium loan quantity $k_{j,t}^{*}$ is:

$$k_{j,t}^{*} = \frac{1}{1-\alpha}\left[1 - \frac{(1-\frac{1-\alpha}{N})(R_t+\tau_j)}{(R_t+\bar{\tau})}\right]k_t^{*} = ms_{j,t}^{*}k_t^{*} \tag{3.18}$$

where $ms_{j,t}^{*} \equiv \frac{1}{1-\alpha}\left[1 - \frac{(1-\frac{1-\alpha}{N})(R_t+\tau_j)}{(R_t+\bar{\tau})}\right]$ denotes the equilibrium market share. If banks are identical, so $\tau_j = \bar{\tau}\ \forall j$, then each bank has a market share of $\frac{1}{N}$ in the Cournot equilibrium. It can be seen that the equilibrium market share depends on the marginal intermediation cost τ_j. More specifically, when bank j has a below average marginal intermediation cost (i.e., $\tau_j < \bar{\tau}$), its market share will be larger than $\frac{1}{N}$. Since $\sum_{j=1}^{N} ms_{j,t} = 1$, $ms_{j,t}^{*}$ must be less than or equal to one given each bank's market share is positive under the parameter restriction on τ_j (3.17). Using (3.17) and (3.18), the fact that $0 < ms_{j,t}^{*} \leqslant 1$ implies that the marginal cost for loans, $R_t + \tau_j$, must lie within the following range:

$$\frac{\alpha(R_t+\bar{\tau})}{(1-\frac{1-\alpha}{N})} \leqslant R_t + \tau_j < \frac{R_t+\bar{\tau}}{(1-\frac{1-\alpha}{N})} \tag{3.19}$$

Proposition 2 which is derived in Appendix B.2.4, shows the bank-specific marginal intermediation cost τ_j affects the extent to which a bank's market share is decreasing in

N.

Proposition 2: *A higher number of banks N reduces the market share of each bank, and this effect is stronger for less efficient banks (with higher τ_j):* $\frac{dms^*_{j,t}}{dN} = -\frac{(R_t+\tau_j)}{N^2(R_t+\bar{\tau})} < 0$.

When banks have different efficiency levels, how bank j's equilibrium loan quantity $k^*_{j,t}$ changes with an increase in the number of banks N depends on the balance between the effect of an increasing aggregate loan quantity and the effect of a falling market share:

$$\frac{dk^*_{j,t}}{dN} = ms^*_{j,t}\frac{dk^*_t}{dN} + k^*_t\frac{dms^*_{j,t}}{dN} \tag{3.20}$$

As N increases (i.e., more intense banking competition), the aggregate loan quantity k^*_t is higher ($\frac{dk^*_t}{dN} > 0$ by Proposition 1), but each bank has a smaller share of the market ($\frac{dms^*_{j,t}}{dN} < 0$ by Proposition 2). Consequently, the sign of $\frac{dk^*_{j,t}}{dN}$ is ambiguous. If the fall in market share of bank j dominates the effect from the increase in total loan quantity, then bank j's loan quantity decreases in N. This requires banks to be identical or have sufficiently similar efficiency levels,[19] as is summarized in Proposition 3, which is proven in Appendix B.2.5.

Proposition 3: *When banks have sufficiently similar efficiency levels such that* $\frac{R_t+\bar{\tau}}{(2-\alpha)(1-\frac{1-\alpha}{N})} < R_t + \tau_j < \frac{R_t+\bar{\tau}}{1-\frac{1-\alpha}{N}}$*, bank j's equilibrium loan quantity $k^*_{j,t}$ unambiguously decreases with N. This condition is satisfied if all banks are identical for $N > 1$.*

Proposition 3 is important for the short-run equity ratio effect which predicts that less banking competition can jeopardize financial stability in the short run. When a reduction in N in period t increases $k_{j,t}$, it leads to a lower equity ratio $\frac{n_{j,t}}{k_{j,t}}$ as the bank's equity $n_{j,t}$ is not affected in period t. Thus, the short-run equity ratio effect operates via the denominator $k_{j,t}$.

In contrast, the long-run equity ratio effect operates via the numerator as competition affects bank's net profit and hence equity accumulated over time, as described in the following proposition, which is derived in Appendix B.2.6.

Proposition 4: *The expected profit of bank j decreases with the number of banks N, as the higher loan rate $R^*_{b,t}$ resulting from less banking competition dominates the changes in loan quantity $k_{j,t}$.*

According to the dynamics of bank equity accumulation (3.9), a higher net profit $\pi^B_{j,t+1}$ leads to a higher $n_{j,t+1}$ and a larger change in bank equity, as long as not all positive profits are distributed as dividends. Together with Proposition 4, this implies

[19]This condition is satisfied for the calibration in Section 3.3.

the long-run equity ratio effect – with less banking competition, banks make higher profits and can accumulate equity faster, leading to higher equity ratios and thereby lower default probabilities over time, as shown next.

Equity Ratio Effect and Banks' Default Probabilities

In this paper, financial stability is measured through banks' default probability $\Gamma(\bar{\epsilon}_{j,t+1})$. By showing how banks' default threshold $\bar{\epsilon}_{j,t+1}$ is determined, the short-run and long-run equity ratio effects on financial stability are explained and compared with the static margin effect.

Since banks cannot diversify away the aggregate risk, an adverse aggregate productivity shock ϵ_{t+1} will cause more entrepreneurs than expected to default and as a result, banks can make negative net profits $\pi^B_{j,t+1}$. If bank j's loss is too large to be absorbed by its equity $n_{j,t}$, its pre-dividend equity $n_{j,t} + \pi^B_{j,t+1}$ will turn negative and it has to default on its liabilities. The threshold for the realized aggregate shock $\bar{\epsilon}_{j,t+1}$ below which bank j defaults is determined by the following condition:

$$\pi^B_{j,t+1}(\bar{\epsilon}_{j,t+1}) + n_{j,t} = 0 \tag{3.21}$$

where $\pi^B_{j,t+1}(\bar{\epsilon}_{j,t+1}) \equiv G(\bar{\epsilon}_{j,t+1})R^*_{b,t}k^*_{j,t} - R_t(k^*_{j,t} - n_{j,t}) - \tau_j k^*_{j,t} - n_{j,t}$ represents the equilibrium net profit when the aggregate shock takes a value of $\bar{\epsilon}_{j,t+1}$, based on (3.8). The LHS of (3.21) represents the pre-dividend equity in period $t+1$. Although the aggregate shock is common to all banks, each bank j's default threshold $\bar{\epsilon}_{j,t+1}$ differs due to their specific τ_j and $n_{j,t}$. Condition (3.21) shows that when the realized aggregate shock takes a value of $\bar{\epsilon}_{j,t+1}$, the proportion of non-performing loans is at such a level that the negative profit for bank j is just absorbed by $n_{j,t}$. If ϵ_{t+1} is below $\bar{\epsilon}_{j,t+1}$, the pre-dividend equity will be negative and bank j will default. Dividing (3.21) by $k^*_{j,t}$ and substituting the definition of $\pi^B_{j,t+1}(\bar{\epsilon}_{j,t+1})$, bank j's default threshold is determined by the following condition:

$$R^*_{b,t}G(\bar{\epsilon}_{j,t+1}) - (R_t + \tau_j) + R_t \frac{n_{j,t}}{k^*_{j,t}} = 0 \tag{3.22}$$

where $R^*_{b,t}G(\bar{\epsilon}_{j,t+1}) - (R_t + \tau_j)$ is the bank's profit margin when the realized aggregate shock takes a value of $\bar{\epsilon}_{j,t+1}$. The bank's revenue fraction $G(\epsilon_{t+1})$ is increasing in the aggregate productivity shock ϵ_{t+1}, as fewer entrepreneurs default, so $G'(\bar{\epsilon}_{j,t+1}) > 0$ as shown in Appendix B.2.7. Let $\kappa_{j,t} \equiv \frac{n_{j,t}}{k^*_{j,t}}$ denote bank j's equilibrium equity ratio. Then it is straightforward to see from (3.22) that banks' default thresholds $\bar{\epsilon}_{j,t+1}$ and hence default probabilities are negatively correlated with their equity ratios. Intuitively, this is because with higher equity ratios, banks can still survive even with a lower realized aggregate shock. The result is summarized in the following proposition, which is formally derived in Appendix B.2.7:

Proposition 5: *Banks' default thresholds $\bar{\epsilon}_{j,t+1}$ are negatively related to banks' equity ratios $\kappa_{j,t}$:* $\frac{d\bar{\epsilon}_{j,t+1}}{d\kappa_{j,t}} = -\frac{R_t}{R^*_{b,t}G'(\bar{\epsilon}_{j,t+1})} < 0 \ \ \forall j$.

Recall that $\Gamma(\epsilon)$ denotes the c.d.f. of the aggregate shock and bank j defaults if $\epsilon_{t+1} < \bar{\epsilon}_{j,t+1}$. So a high default threshold $\bar{\epsilon}_{j,t+1}$ leads to a high default probability $\Gamma(\bar{\epsilon}_{j,t+1})$. Thus, Proposition 5 implies a negative relationship between banks' default probabilities and their equity ratios.

This paper focuses on how imperfect banking competition affects banks' equity ratio and hence their default probabilities. The role of the equity ratio effects can be shown by implicitly differentiating bank j's default condition (3.21) with respect to the number of banks N, as shown in Appendix B.2.8:

$$\frac{d\bar{\epsilon}_{j,t+1}}{dN} = \frac{\overbrace{R_t \frac{n_{j,t}}{k^*_{j,t}} \frac{dk^*_{j,t}}{dN} \frac{1}{k^*_{j,t}}}^{\text{SR equity ratio effect}} \quad \overbrace{-R_t \frac{1}{k^*_{j,t}} \frac{dn_{j,t}}{dN}}^{\text{LR equity ratio effect}} \quad \overbrace{-\frac{dR^*_{b,t}}{dN} G(\bar{\epsilon}_{j,t+1})}^{\text{margin effect}}}{R^*_{b,t}G'(\bar{\epsilon}_{j,t+1})} \tag{3.23}$$

Suppose that banks' efficiency levels are sufficiently similar so that $\frac{dk_{j,t}*}{dN} < 0$ by Proposition 3. According to (3.23), when N is lower, there is a short-run equity ratio effect that predicts a higher default probability due to a lower equity ratio, provided that the bank has equity $n_{j,t} > 0$. This is because when N is lower, each bank has greater market power and hence a larger loan quantity $k^*_{j,t}$. This reduces bank j's equity ratio $\frac{n_{j,t}}{k^*_{j,t}}$ for a given $n_{j,t}$, which leads to a higher default threshold $\bar{\epsilon}_{j,t+1}$. In addition, there is a long-run equity ratio effect such that a lower N tends to raise future equity via higher profits (by Proposition 4), which increases bank j's equity ratio in the long run and thereby reduces its future default threshold. In contrast, the static margin effect predicts that a lower N reduces the default threshold $\bar{\epsilon}_{j,t+1}$ due to a higher loan rate (as $\frac{dR^*_{b,t}}{dN} < 0$ by Proposition 1) and thus higher revenue from performing loans, which provide a buffer against loan losses.

How $\bar{\epsilon}_{j,t+1}$ changes with N in the short run (when $\frac{dn_{j,t}}{dN} = 0$) is ambiguous theoretically. For the calibration of the model described in the next section, the short-run equity ratio effect tends to dominate the static margin effect and as a result, less banking competition tends to raise banks' default probabilities and undermine financial stability in the short run. Over time, the short-run equity ratio effect tends to disappear, provided that banks retain their profits to build up equity. These results are summarized in Proposition 6, which is formally shown in Appendix B.2.8.

Proposition 6: *In the short run, less banking competition can jeopardize financial stability by lowering banks' equity ratios. However, when banks retain the greater profits to build up their equity over time, less banking competition can enhance financial stability.*

The extent of the financial stability gain from less banking competition over time (long-run equity ratio effect) depends on banks' dividend distribution or macroprudential policies. For instance, if banks do not distribute to shareholders and equity is accumulated over time via past profits, this can lead to a large gain in financial stability from less banking competition, which is shown in Section 3.4. By restricting banks' dividend payment to shareholders, macroprudential policies can thus help to ensure a larger gain in financial stability under imperfect banking competition.

The model is first calibrated in Section 3.3 and then simulated to illustrate three model implications. First, less banking competition can lead to a large gain in financial stability provided that banks accumulate equity over time. This shows the relevance of macroprudential regulation on banks' dividend distribution. Second, a bank merger that reduces banking competition can raise the default probability of the merged bank by lowering its equity ratio. Third, less banking competition leads to a larger macroeconomic efficiency loss and whether this efficiency loss outweighs the financial stability gain depends on the extent of equity accumulation and macroeconomic volatility.

3.3 Calibration

The model is calibrated to match the data for Germany during the period of 1999-2014 on nine key variables, i.e., 5-bank asset concentration ratio, HHI concentration index (Herfindahl-Hirschman Index), mean market share, corporate lending rate, interest income to total assets ratio, non-interest expense to total assets ratio, loan impairment cost ratio, equity ratio, and bank's default probability. The first four variables indicate banks' market power and concentration in the banking sector, while the remaining variables include ratios that indicate banks' leverage, profitability and cost efficiency.

The last column in Table 3.1 shows the mean values of the nine variables in the data for Germany. The net corporate lending rate ($R_{b,t} - 1$) is empirically constructed by averaging two country-level corporate lending rate series (i.e., for loans of up to 1 million EUR and for loans of over 1 million EUR) across years (2000-2014), where the lending rates are from the ECB. HHI is the sum of squared market shares of all banks, where the market share of a given bank in a given year $\frac{k_{j,t}}{k_t}$ is computed as the ratio of the bank's total assets to the sum of total assets of all banks in that year.[20] HHI ranges from $\frac{1}{N}$ to one and a higher value implies higher bank concentration. The 5-bank concentration ratio is the sum of the market shares of the five largest banks by total assets. Both concentration measures are obtained from the ECB and the numbers reported in Table 3.1 are mean values over the period of 1999-2014. The remaining variables are calculated

[20]Loans are assumed to be the only assets on banks' balance sheets in the model, so total assets are used to proxy for $k_{j,t}$ empirically.

using Bankscope annual balance sheet data for six types of banks in Germany during the period of 1999-2014.[21]

The loan impairment cost ratio in the model is $\frac{R_{b,t}k_{j,t}(1-\mathrm{E}_t[G(\epsilon_{t+1})])}{R_{b,t}k_{j,t}}$, where the numerator reflects the loss in gross loan revenue due to non-performing loans and the denominator is the gross loan revenue if all loans are repaid. Empirically, the loan impairment charge is used to proxy for the numerator and gross loans are used as the denominator since gross loan revenue is not available in data. So the loan impairment charge to gross loans ratio is used to proxy for the loan impairment cost ratio $(1 - \mathrm{E}_t[G(\epsilon_{t+1})])$ in the model. The average total equity to total assets ratio $\frac{n_{j,t}}{k_{j,t}}$ across banks in Germany over 1999-2014 is around 7.2%, so the desired equity ratio κ^* is set to be 7.2%. Interest income to total assets $\frac{\pi^B_{j,t}}{k_{j,t}}$ is calculated as the gross interest and dividend income net of the total interest expense over total assets. The marginal intermediation cost τ_j in the model is empirically proxied by total non-interest expenses to total assets ratio. Bank's default probability in the model is $\Gamma(\bar{\epsilon}_{j,t+1})$, where $\Gamma(.)$ is the c.d.f. the aggregate shock distribution. The risk-neutral annual bank's default probability of 2.01% in the data is computed from the average CDS spread of 122 basis points across German banks during 2003-2014 based on Hull (2012).[22]

Table 3.1: Matching Key Variables with Data for Germany During 1999-2014

Variable	Model (N=60) Identical τ	Model (N=60) Heterogeneous τ	Data Germany
5-bank asset concentration	0.083	0.229	0.249
HHI (total assets)	0.017	0.025	0.021
Net corporate lending rate	5.07%	5.07%	4.06%
Loan impairment charge/gross loans	0.006	0.006	0.006
Non-interest expense/total assets	0.032	0.032	0.026
Bank's default probability	2.13%	2.13%	2.01%
Interest income/total assets	0.012	0.012	0.024
Mean market share	0.017	0.017	0.000
Desired total equity/total assets	0.072	0.072	0.072

Data sources: ECB, Bankscope, Thomson Reuters EIKON
Note: The numbers reported in the last column are mean values across banks and years (across years) for bank-level (country-level) variables. Data on the first three variables are from the ECB. The remaining variables except for bank's default probability are computed using Bankscope annual statements. Variables from Bankscope are winsorized at 1% of the top and the bottom of the distribution. Bank's default probability is calculated using the average CDS spread across banks from Thomson Reuters EIKON. All the numbers in the two model columns are model results, except for the desired equity to assets ratio which is calibrated.

[21] The sample of banks used consists of bank holding companies, commercial banks, cooperative banks, finance companies, real estate & mortgage banks, and savings banks.

[22] Following Egan et al. (2017) and Hull (2012), the probability of default is calculated under a risk neutral model with a constant hazard rate, assuming that the recovery rate is 40% and the risk-free rate or LIBOR is 3%.

I calibrate the model parameters to match the value of these nine variables in the data for Germany. Table 3.1 compares the mean values of these variables computed using the simulated data with those in the real data. The capital share α is set at 0.3. A value of α larger than 0.5 leads to an unrealistically large gross loan rate.[23] The desired equity-to-assets ratio is set at 0.072 to match the average equity ratio of around 7.2% across banks in Germany during 1999-2014. Both the deterministic productivity level A and the exogenous gross deposit rate R is set at one. The distribution for the idiosyncratic productivity shock ω is assumed to be lognormal with the mean set to be -0.15, so that the probability of entrepreneurs' default $F(\bar{\omega}_{t+1})$ is around 3%, following the literature (e.g., Bernanke et al., 1999). Since the expected value of ω is one, the variance of the log-normal distribution needs to be 0.3. The distribution for the aggregate productivity shock ϵ is assumed to be lognormal with the variance chosen to be 0.28 to match the average default probability of 2.01% in data. The collection cost parameter $\mu \in [0, 1]$ is set at 0.04 to match the mean loan impairment charge to gross loans ratio of 0.006.

Banks' marginal intermediation costs τ_j are randomly drawn from a reverse bounded Pareto distribution with a support of $[0.001, 0.04]$ and a shape parameter of 0.1.[24] The distribution needs to be bounded to ensure a non-negative market share. In essence, the bounded Pareto distribution is the conditional distribution that results from restricting the domain of the Pareto distribution to $[0.001, 0.04]$. As shown in (3.17), τ_j cannot be larger than a certain factor of the mean $\bar{\tau}$, otherwise bank j is too inefficient to operate profitably. The bounded Pareto distribution is reversed to give a long left tail, such that most simulated banks have a τ_j close to 0.04 and only a few banks will have a relatively low τ_j close to 0.001, resulting in a market share distribution with a few large dominant banks and a lot of small inefficient banks. High bank concentration and a small mean market share across banks in the data, as can be seen in Table 3.1, indicate that the banking sector tends to be dominated by a few dominant players with high market shares, alongside a lot of small banks with very low market shares. The support of $[0.001, 0.04]$ is chosen to match the average non-interest expense to total assets ratio (or $\bar{\tau}$ in the model) of 0.026. The shape parameter is chosen to be 0.1 to give a skewed distribution for market shares. More details on the distribution for τ can be found in Appendix B.3.1.

Together with the calibration for the aggregate shock distribution, the number of banks N is chosen to be 60 to match the concentration measures, mean market share, average corporate lending rate, and the interest income to assets ratio with empirical data. Further increasing N lowers the mean market share and brings it closer to the

[23]For example, $\alpha = 0.7$ gives R_b higher than 1.5 for N ranging from 1 to 20.

[24]Increasing the upper bound of the support will raise the mean marginal intermediation cost across banks $\bar{\tau}$ and the equilibrium loan rate. The support of $[0.001, 0.04]$ means the lowest and highest value that τ_j can take is 0.001 and 0.04 respectively. The shape parameter is the tail index and a smaller value gives a heavier tail. The Pareto distribution is a skewed and heavy-tailed distribution that allows a more dispersed distribution of bank efficiency, which gives rise to a more unequal market share distribution.

mean market share of almost zero in data, but at the same time, it also reduces the interest income to total assets ratio due to a lower equilibrium loan rate.

Table 3.1 shows that bank concentration measures computed using the simulated data based on the model with heterogeneous banks (with different τ_j) are very close to the measures in the data. In contrast, the concentration measures predicted by the model with identical banks ($\tau_j = \bar{\tau} \ \forall j$) are much lower than those in the data. This is because with $N = 60$, each identical bank only has a small market share. A summary table of the calibrated parameters is shown in Table B.1 in Appendix B.3.2.

3.4 Simulation Results

Using the calibrated model, this section illustrates the long-run equity ratio effect, the short-run equity ratio effect and quantifies the relative importance of the financial stability gains and macroeconomic efficiency losses associated with imperfect banking competition.[25] Section 3.4.1 shows the average financial stability gain across heterogeneous banks relative to the perfect banking competition benchmark over time, for different numbers of banks N and different bank dividend distribution or macroprudential policies. In addition, it also shows the financial stability gain for banks with different market shares relative to the perfect banking competition benchmark over time under a given level of banking competition (or a given N). Section 3.4.2 uses a bank merger scenario to illustrate the short-run equity ratio effect. Section 3.4.3 quantifies the macroeconomic efficiency loss from imperfect banking competition and constructs a new measure to compare its importance relative to the financial stability gain.

3.4.1 Financial Stability Gain from Imperfect Banking Competition

This section illustrates the long-run equity ratio effect, assuming all banks have the same initial equity ratio across different levels of banking competition, to focus on the effect of the number of banks N on financial stability.

In this section, the financial stability gain of bank j is measured by the difference between the default probability of the representative bank under perfect banking competition $\Gamma(\bar{\epsilon}^{PC}_{t+1})$ and bank j's default probability under imperfect banking competition $\Gamma(\bar{\epsilon}_{j,t+1})$:

$$\text{Financial Stability Gain of Bank } j = \Gamma(\bar{\epsilon}^{PC}_{t+1}) - \Gamma(\bar{\epsilon}_{j,t+1}) \tag{3.24}$$

[25]The model consists of a few systems of nonlinear equations, which are solved using Julia JuMP and Ipopt. More specifically, after solving for the equilibrium loan rate and the market shares, the profit of each bank is known and the equity accumulation process can be determined for each bank under different banks' dividend distribution or macroprudential policies. Given the equity dynamics of banks, their default thresholds or default probabilities in each period can be solved.

where $\Gamma(\epsilon)$ is the continuous c.d.f. for the aggregate shock ϵ. Following (3.22), the default threshold of the representative bank $\bar{\epsilon}_{t+1}^{PC}$ is determined by:

$$R_{b,t}^{PC} G(\bar{\epsilon}_{t+1}^{PC}) - (R_t + \bar{\tau}) + R_t \frac{n_t}{k_t} = 0 \tag{3.25}$$

where the equilibrium gross loan rate $R_{b,t}^{PC}$ under perfect competition can be found from the equilibrium loan rate (3.15) by setting N to infinity. The representative bank is assumed to have a marginal intermediation cost of $\bar{\tau}$, which equals the mean marginal intermediation cost across banks under imperfect banking competition. Without aggregate shocks, the representative bank under perfect competition always makes a zero profit, so the equity ratio $\frac{n_t}{k_t}$ is equivalent to its initial level and its default threshold $\bar{\epsilon}_{t+1}^{PC}$ is constant over time. By contrast, each bank under imperfect competition has a different marginal intermediation cost and hence a different profit margin, which leads to differences in equity ratios and thus default thresholds across banks and over time.

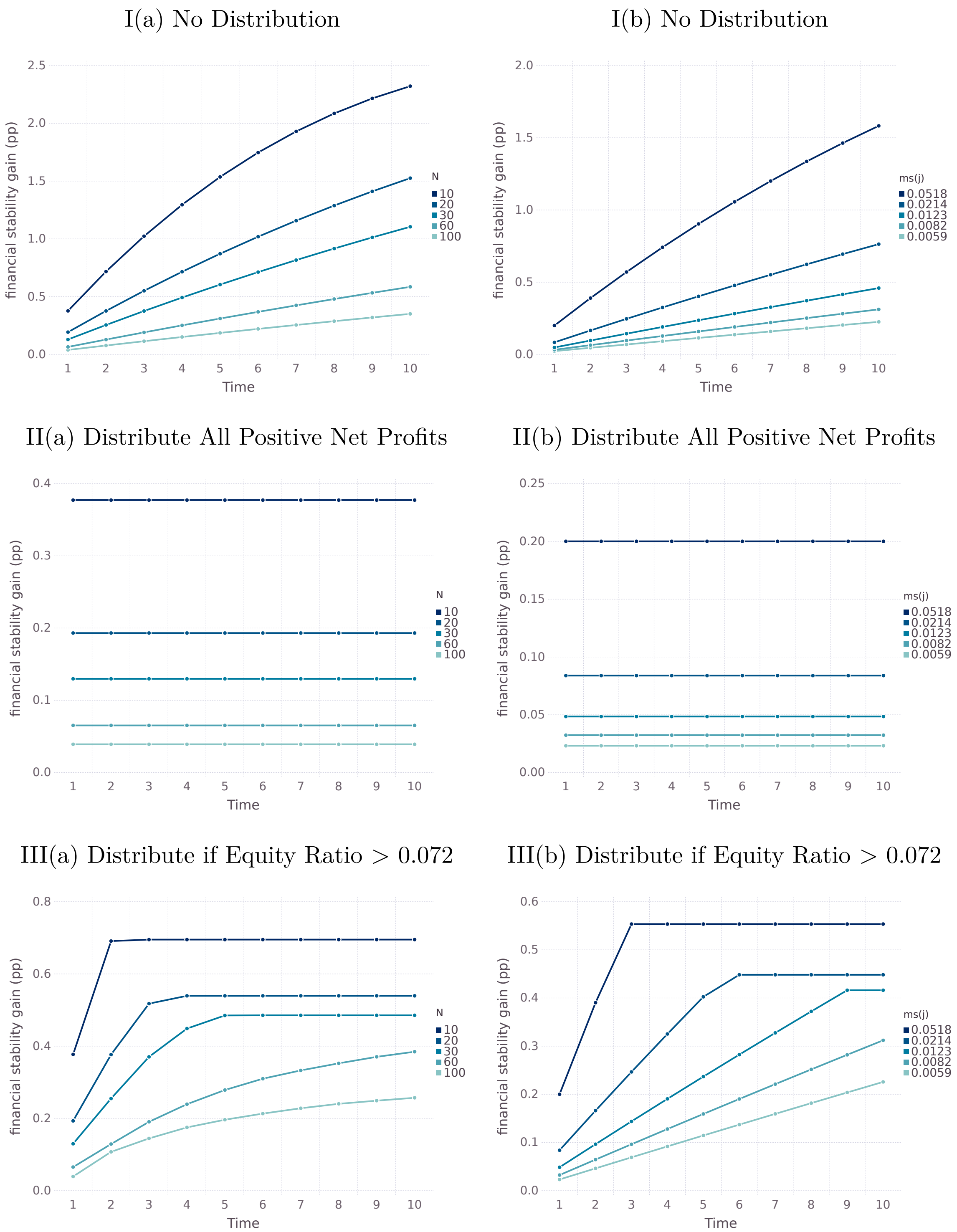
Figure 3.2 plots the average financial stability gain across heterogeneous banks for different levels of competition (i.e., different number of banks N) and the financial stability gain of banks with different market shares for a given (baseline) level of N.[26] In each case, the effects on financial stability gains over time are shown for the three different bank dividend distribution or macroprudential policies presented in Section 3.2.2. For each N, banks' marginal intermediation costs τ_j are randomly drawn from the same reverse bounded Pareto distribution. To focus on the effects coming from imperfect banking competition for a fixed N, assume the realization of aggregate shocks is $\epsilon = 1$ throughout.

Graphs in the first column of Figure 3.2 plot the average financial stability gain (in percentage points) of heterogeneous banks under imperfect banking competition, so $\frac{1}{N}\sum_j \left(\Gamma(\bar{\epsilon}_{t+1}^{PC}) - \Gamma(\bar{\epsilon}_{j,t+1})\right) * 100$, following (3.24). As can be seen in graph I(a) of Figure 3.2, the average financial stability gain across banks in period 1 is slightly higher for smaller N (i.e., less banking competition), which is purely caused by the static margin effect as banks are assumed to start with the same initial equity ratio for simplicity. But the differences in financial stability gain across different N are amplified over time due to bank equity accumulation that leads to higher equity ratios over time. By contrast, if all positive net profits are distributed away, as shown in graph II(a), bank equity accumulation is absent and hence the financial stability gain does not increase over time. The differences in the mean financial stability gain across banks for different N are only caused by the static margin effect in this case.

Graph III(a) shows the case where banks only distribute profits if their equity ratios exceed the desired or required level κ^*, which is calibrated to be 0.072 based on the average equity ratio across banks in Germany during 1999-2014. Starting with zero initial equity, for smaller values of N, banks face less competition, so they have higher

[26]The mean financial stability gain of identical banks with the same marginal intermediation cost $\bar{\tau}$ for different N gives very similar results to the case of heterogeneous banks with different τ_j, so the former case is not shown in this section.

Figure 3.2: Financial Stability Gain from Imperfect Banking Competition



Note: Financial stability gain is measured by the differences in banks' default probabilities between perfect and imperfect banking competition, based on (3.24). Graphs in the first column plot the average stability gain (in percentage points) across heterogeneous banks with different marginal intermediation cost τ_j over time for different numbers of banks N, while graphs in the second column plot the stability gain (in percentage points) of 5 different banks at 5 percentiles (1st, 25th, 50th, 75th, 99th) of the equilibrium market share ms_j^* for $N = 60$. Each row shows a different case of bank dividend distribution or macroprudential policies.

profits and accumulate equity faster, resulting in lower default probabilities compared to the perfect banking competition benchmark. As a result, their financial stability gains increases more quickly during the first few periods. Once banks' equity ratios reach κ^*, (positive) profits are distributed to shareholders, so there is no further increase in financial stability gains, as shown in graph III(a).

Comparing graph I(a) without dividend distribution to graph II(a) with full distribution of positive profits shows the power of the long-run equity ratio effect. Since the financial stability gain from less banking competition is largely attributed to the accumulation of greater profits over time rather than the static margin effect, macroprudential policies that limit banks' dividend distribution can significantly increase the financial stability gains.

Graphs in the second column of Figure 3.2 plot the financial stability gain of banks with different market shares under a given level of competition ($N = 60$ as in the baseline calibration) over time. Banks are ranked according to their marginal intermediation costs τ_j, where banks with lower τ_j have higher equilibrium market shares. Five banks at five different percentiles (1st, 25th, 50th, 75th, 99th) of τ_j are plotted. The legend shows the corresponding equilibrium market share for each bank ms_j. As can be seen in graph I(b), when there is no dividend distribution, a more efficient larger bank has a higher financial stability gain over time due to a higher profit margin and faster equity accumulation. By contrast, when profits are distributed away, as shown in graph II(b), differences in financial stability gain between banks are purely caused by the differences in profit margins (margin effect), which are relatively small compared to the differences caused by equity accumulation. Graph III(b) shows that when banks distribute only if their equity ratios exceed κ^*, the financial stability gain of a larger bank is higher due to a larger profit margin and faster equity accumulation. Once banks' equity ratios reach κ^*, the differences in the financial stability gain across banks are purely caused by the margin effect. As shown in graph III(b), the smallest bank with a market share of 0.59% accumulates equity very slowly due to a lower profit margin and even after 10 periods, its equity ratio still has not reached κ^*.

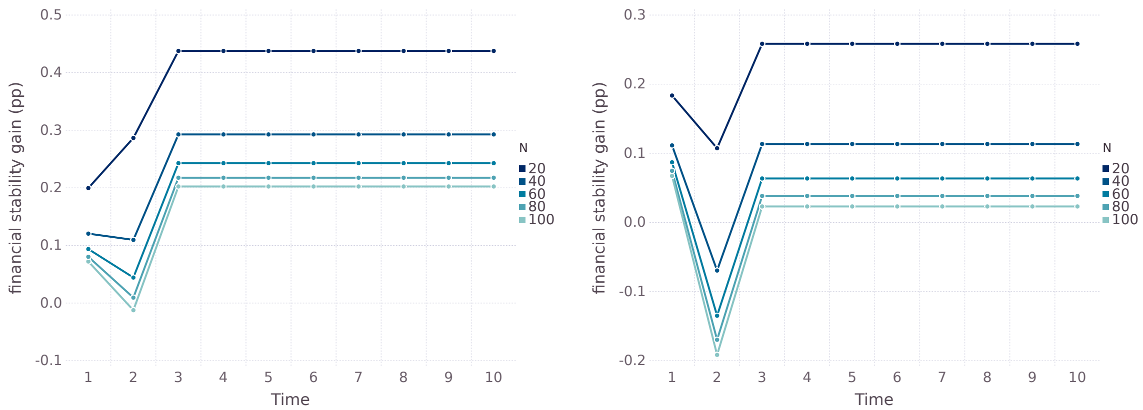
3.4.2 Bank Merger Scenario

This section illustrates the short-run equity ratio effect using a bank merger scenario where solvent banks that survived a crisis merge with distressed banks that have little equity. This is an interesting case to look at since the massive public intervention and bank mergers during the 2007-2009 crisis have distorted banking competition and led to increased bank concentration in many countries (Vives, 2011).[27] Unlike an increase in the number of banks N, which is clearly due to new entrants, a reduction in N can be caused

[27] Perotti and Suarez (2002) specifically look at the merger policy that promotes takeovers of failed banks by solvent banks and argue that this policy can reinforce financial stability by raising banks' expected profits and thus reducing their risk taking. However, I find that these bank mergers can

Figure 3.3: Financial Stability Gain after Solvent Banks Merge with Distressed Banks

(a) Distribute if Equity Ratio > 0.072 (b) Distribute if Equity Ratio > 0.109



Note: The two graphs plot the average financial stability gain (in percentage points) across the merged banks, computed using (3.24), under different initial levels of N as shown in the legend. In graph (a) vs (b), half of the banks with an equity ratio of 0.072 vs 0.109 merge in period 1 with the other half distressed banks with zero equity, so N reduces to $\frac{N}{2}$ from period 1 onwards. The desired equity ratio κ^* is assumed to be 0.072 vs 0.109 in graph (a) vs (b).

by bank exits or bank mergers. The magnitude and the direction of the short-run equity ratio effect depend on the causes for the reduction in N. In the case of bank exits, the remaining banks' equity levels are unaffected, so their equity ratios unambiguously fall after the increase in concentration. By contrast, in the case of a bank merger, the equity of the merged bank (which is the sum of the two banks' equity levels before the merger) increases. Nevertheless, when a solvent bank merges with a distressed bank with little equity, the equity of the merged bank may not increase as much relative to the increase in loan quantity due to the greater market power, resulting in a lower equity ratio.[28]

Assume there are N banks in period 0, with half being relatively more efficient with a lower marginal intermediation cost τ_j of 0.024, and the other half being less efficient with a higher marginal intermediation cost of 0.04.[29] An adverse aggregate productivity shock in period 0 wipes out the equity of those inefficient banks, so they are left with zero equity. The efficient banks are also affected by the aggregate shock but are less badly hit and have a positive equity ratio of $\kappa^* = 0.072$ in graph (a) of Figure 3.3. In period 1, each solvent bank merges with one distressed bank whose equity was wiped out in period

undermine financial stability in the short run by looking at their effect on the equity ratios of the merged banks.

[28]In the model, when two identical banks with the same equity ratio merge, the increase in equity will more than offset the increase in loan quantity, leading to an increase in its equity ratio and a lower default probability even in the short run. Technically, each merged bank has greater market power and thus a higher loan quantity after the merger. However, the merged bank's loan quantity does not double compared to each individual bank's loan quantity before the merger due to the fall in aggregate loan quantity with less banking competition, while its equity doubles, so its equity ratio goes up in the short run.

[29]In this case, the inefficiency τ_m of the merged bank is the average of the two banks before the merger, which is 0.032, equivalent to the average bank inefficiency in the baseline calibration shown in Table 3.1.

0.[30] Assume the inefficiency τ_m of the merged bank is the average inefficiency τ_j of the two banks before the merger, so the mean bank inefficiency $\bar{\tau}$ does not change with N.

The bank mergers in period 1 reduce the number of banks to $\frac{N}{2}$ from period 1 onwards, resulting in less banking competition. Consequently, the financial stability gain would be expected to increase due to the higher profit margin on performing loans that provides a buffer against loan losses (margin effect). The graphs in Figure 3.3, however, show the opposite in most cases. The graphs plot the average financial stability gain (in percentage points) of a merged bank after the bank mergers for different initial numbers of banks N before the mergers. As long as N is not too small (e.g., $N = 20$ in graph (a)), the financial stability gain falls after the bank mergers. This is because the short-run equity ratio effect (due to the drop in the merged bank's equity ratio) dominates the margin effect. Since the solvent bank does not inherit much equity from the distressed bank, the equity of the merged bank only increases a little after the merger. Meanwhile, the merged bank has greater market power and thus a larger loan quantity under less banking competition, so the equity ratio of the merged bank falls. The lower equity ratio of the merged bank in period 1 leads to a fall in financial stability gain in period 2. Since the margin effect is weaker with larger N (i.e., more intense banking competition), the fall in financial stability gain is more noticeable for larger N.

Furthermore, when the initial equity of the solvent bank is larger, the short-run equity ratio effect is stronger, as shown in graph (b). This follows from (3.23), which shows that the short-run equity ratio effect is absent when the initial equity is zero. When the initial equity is larger, the increase in loan quantity caused by the merger has a larger impact on reducing the merged bank's equity ratio. As shown in graph (b), where solvent banks are assumed to have an initial equity ratio of 0.109,[31] even when N decreases from 20 to 10 after the bank mergers, the short-run equity ratio effect still dominates the margin effect and hence the financial stability gain falls in period 2. So in the short run, less banking competition can jeopardize financial stability due to a stronger short-run equity ratio effect relative to the static margin effect.

3.4.3 Efficiency Loss from Imperfect Banking Competition

This section quantifies the macroeconomic efficiency loss associated with imperfect banking competition in terms of the reduction in expected output and compares it with the financial stability gain. The macroeconomic efficiency loss from imperfect banking competition is computed as:

$$\text{Macroeconomic Efficiency Loss} = \frac{\mathrm{E}_t(y_{t+1}^{PC}) - \mathrm{E}_t(y_{t+1})}{\mathrm{E}_t(y_{t+1}^{PC})} \tag{3.26}$$

[30] Just before the merger in period 1, however, distressed banks also have some equity due to the realized profits in period 1.

[31] The mean equity ratio across EU banks from 1999-2014 is 0.109.

where $\mathrm{E}_t(y_{t+1}) = A(k_t^*)^\alpha$ is the expected output with imperfect banking competition when the loan rate is $R_{b,t}^*$ (3.15). Based on Proposition 1 in Section 3.2.3, the macroeconomic efficiency loss is larger with less banking competition due to a higher loan rate and thus a lower demand for physical capital and lower expected output.

To compare the financial stability gain with the macroeconomic efficiency loss, this section constructs a new measure of financial stability gain in real terms. The financial stability loss of a bank j is the part of the liabilities (deposits) that the bank defaults on when it goes bankrupt. More specifically, when the realized aggregate productivity shock is sufficiently low, i.e., $\epsilon_{t+1} < \bar{\epsilon}_{j,t+1}$, bank j's loss (or negative net profit $\pi_{j,t+1}^B$) is too large to be absorbed by its equity $n_{j,t}$, so $\pi_{j,t+1}^B(\epsilon_{t+1}) + n_{j,t}$ represents the unabsorbed loss of bank j or the amount of liabilities (deposits) that bank j defaults on. Since depositors are assumed to be protected by a full deposit guarantee, in this case, the government steps in to repay the bank's depositors. So the financial stability loss when bank j goes bankrupt is $\int_0^{\bar{\epsilon}_{j,t+1}} \left(\pi_{j,t+1}^B(\epsilon) + n_{j,t}\right) d\Gamma(\epsilon)$, which is negative.

With perfect banking competition, the representative bank is more likely to default due to a lower profit margin and a lower equity ratio over time, so the expected loss in financial stability when the representative bank defaults is even larger. Hence, there is a financial stability gain of bank j from imperfect banking competition relative to perfect banking competition. The financial stability gain from imperfect banking competition normalized by the expected output under perfect banking competition is constructed as follows:

$$\text{Financial Stability Gain} = \frac{\sum_j \int_0^{\bar{\epsilon}_{j,t+1}} \left(\pi_{j,t+1}^B(\epsilon) + n_{j,t}\right) d\Gamma(\epsilon) - \int_0^{\bar{\epsilon}_{t+1}^{PC}} \left(\pi_{t+1}^B(\epsilon) + n_t\right) d\Gamma(\epsilon)}{\mathrm{E}_t(y_{t+1}^{PC})} \tag{3.27}$$

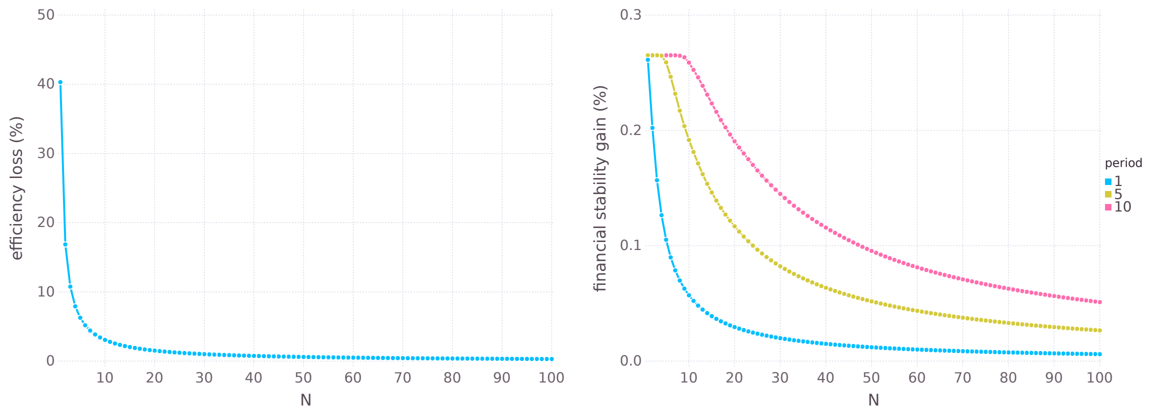
where $\Gamma(\epsilon)$ is the c.d.f. of the aggregate shock distribution and $\bar{\epsilon}_{t+1}^{PC}$, π_{t+1}^B and n_t represent the default threshold, net profit and equity of the representative bank under perfect banking competition respectively. The default threshold $\bar{\epsilon}_{t+1}^{PC}$ is calculated using the same method as shown in (3.25) in Section 3.4.1. The first term in the numerator of (3.27) represents the total financial stability loss of banks from imperfect banking competition and the second term shows the financial stability loss of the representative bank from perfect banking competition.

To quantify the importance of the financial stability gain relative to the macroeconomic efficiency loss from imperfect banking competition, I construct the following net gain measure:

$$\begin{aligned}\text{Net Gain} &= \text{Financial Stability Gain} - \text{Macroeconomic Efficiency Loss} \\ &= \frac{\sum_j \int_0^{\bar{\epsilon}_{j,t+1}} (\pi_{j,t+1}^B(\epsilon) + n_{j,t}) d\Gamma(\epsilon) - \int_0^{\bar{\epsilon}_{t+1}^{PC}} (\pi_{t+1}^B(\epsilon) + n_t) d\Gamma(\epsilon) - [\mathrm{E}_t(y_{t+1}^{PC}) - \mathrm{E}_t(y_{t+1})]}{\mathrm{E}_t(y_{t+1}^{PC})}\end{aligned} \tag{3.28}$$

Figure 3.4: Macroeconomic Efficiency Loss and Financial Stability Gain

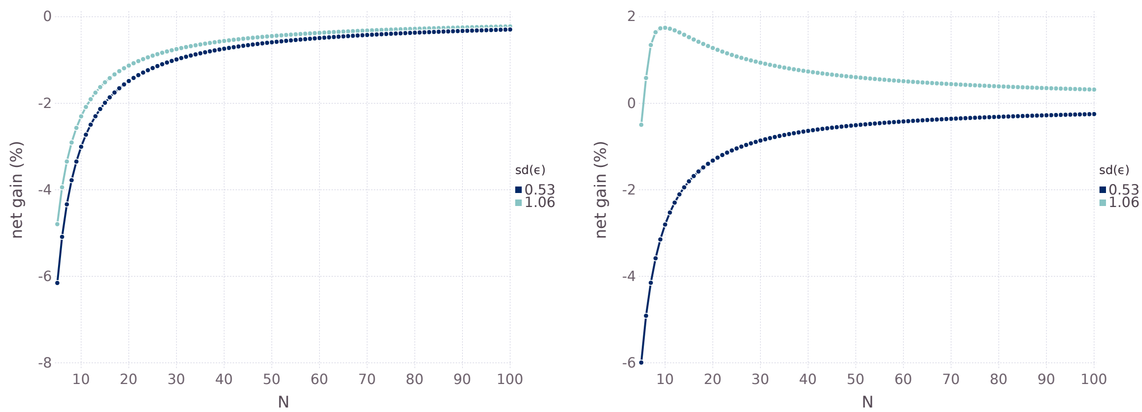
(a) Macroeconomic Efficiency Loss (b) Financial Stability Gain



Note: Graph (a) plots the macroeconomic efficiency loss (%) based on (3.26) across different levels of banking competition, with the number of banks N ranging from 1 to 100. Assuming there is no dividend distribution to shareholders, graph (b) plots the output measure for financial stability gain (%) based on (3.27) in period 1, 5, and 10 respectively, with the baseline calibration for the standard deviation of the aggregate shock ϵ distribution (i.e., sd(ϵ) = 0.53) and different N ranging from 1 to 100.

Figure 3.5: Compare Macroeconomic Efficiency Loss with Financial Stability Gain

(a) Net Gain in Period 1 (b) Net Gain in Period 10



Note: Graphs (a) and (b) plot the net gain (%) based on (3.28) in period 1 and 10 respectively, for different number of banks N ranging from 5 to 100 and different standard deviations of the aggregate shock ϵ distribution, assuming there is no dividend distribution to shareholders.

which is the difference between the financial stability gain (3.27) and the macroeconomic efficiency loss (3.26) resulting from imperfect banking competition. As can be seen from (3.28), the net gain measure is positive when the financial stability gain outweighs the macroeconomic efficiency loss.

Graph (a) in Figure 3.4 plots the macroeconomic efficiency loss across different number of banks N, which is computed following (3.26). As can be seen in graph (a), there is a large macroeconomic efficiency loss when there is very little competition (i.e., N is very small). For example, with a monopoly bank, the expected output is 40% lower than that with a perfectly competitive banking sector. When N increases, the loan rate becomes lower and approaches the the loan rate under perfect banking competition, which leads to a higher demand for physical capital and higher expected output.

Graph (b) in Figure 3.4 plots the output measure for the financial stability gain across different number of banks N and in three different time periods, which is computed following (3.27). The graph is plotted under the baseline calibration for the standard deviation of the aggregate shock distribution of 0.53, which gives a bank default probability of around 2.13%. As the standard deviation increases during the volatile times, for instance, the financial stability gain also rises. As can be seen from graph (b), the financial stability gain increases over time due to bank equity accumulation under imperfect banking competition, which leads to higher bank equity ratios and thus lower bank default probabilities.

Graphs (a) and (b) in Figure 3.5 plot the net gain measure based on (3.28) with different number of banks N and different standard deviations of the distribution for the aggregate shock ϵ in period 1 and 10 respectively.[32] Assume banks are identical and have zero initial equity across different levels of N (including perfect banking competition when N approaches infinity). The differences in the average financial stability gain across different N in period 1 are caused by the margin effect. As can be seen from graph (a), when there is only the static margin effect, the net gain measure is negative and approaches zero as N tends to infinity. This is because in the absence of bank equity accumulation, the financial stability gain from imperfect banking competition is very small and is always outweighed by the macroeconomic efficiency loss. In this case, perfect banking competition is the best.

However, as banks under imperfect banking competition accumulate equity over time and have higher equity ratios than their counterparts under perfect banking competition, the net gain starts to turn positive during more volatile times when the standard deviation of the aggregate shock distribution is high, implying that the financial stability gain can outweigh the macroeconomic efficiency loss over time, as shown in graph (b). This also depends on the degree of imperfect banking competition.

More specifically, when there is very little competition (i.e., the number of banks is below 5), the macroeconomic efficiency loss is very large, as shown in graph (a) in Figure 3.4, which overrides any financial stability gain. As a result, even in the presence of bank equity accumulation, the net gain is still negative for small values of N. When there are more than six banks, the macroeconomic efficiency loss from imperfect banking competition is not too substantial and the financial stability gain due to equity accumulation over time can outweigh the macroeconomic efficiency loss, as can be seen from graph (b) in Figure 3.5. The net gain in period 10 is almost 2% when there are ten banks and aggregate volatility is high.

[32] In Figure 3.5, the number of banks N ranges from 5 to 100 to make the differences between the lines more noticeable.

3.5 Data

Bank-level data on annual financial statements information are from Bankscope, which are used to calculate national concentration in the banking sector (i.e., Herfindahl Hirschman Index (HHI) and the 5-bank asset concentration ratio). The ECB Macroprudential database provides these two concentration measures estimated based on the total assets of credit institutions authorized in a given country, however, these measures are only available for EU countries. Using Bankscope data, I compute the two concentration measures for both EU and OECD countries.

There are two difficulties in computing the national banking concentration. First, some banks can have multiple statements with different consolidation code in Bankscope (i.e., unconsolidated statements U2 and consolidated statements C2).[33] To avoid double counting, only one of U2 or C2 should be kept for each bank. Keeping C2 means consolidated statements are used wherever possible when computing bank concentration. This may be appropriate if the controlled subsidiaries are domestic, so using consolidated statements may better reflect the national bank concentration. However, if the controlled subsidiaries are foreign, then using consolidated statements can overestimate the national concentration. In this paper, I choose to drop C2 and use unconsolidated statements wherever possible because the resulted measures align more closely with the ECB estimates. I use six different types of banks to compute the concentration, i.e., bank holding companies, commercial banks, cooperative banks, finance companies, real estate & mortgage banks, and savings banks, since this paper focuses on the types of banks whose main business is making loans. The types of banks that are dropped only account for around 5% of the total observations. Graphs for the two concentration measures over time for each EU or OECD countries in Figure B.1 and B.2 in Appendix B.4.3 show that in general, the ECB concentration measures have a smaller magnitude than my own calculation since the sample of banks used by ECB is likely to be larger than my sample.

Second, as noted by Uhde and Heimeshoff (2009), the sample of banks tends to increase over time in Bankscope, so the observed variation in concentration may be caused by the data coverage issue. To avoid this problem, I checked the data coverage for each EU or OECD country in each year using aggregate-level total assets and total credit data from the ECB and Bank for International Settlements (BIS) respectively. More specifically, for each EU country in each year, the sum of total assets of banks from Bankscope is divided by the total assets of all credit institutions from ECB and a larger ratio indicates better coverage. For each OECD country in each year, the sum of gross loans of banks from Bankscope is divided by the total credit of domestic banks (to

[33] There are four main consolidation types in Bankscope, U1, U2, C1, and C2. U2 (U1) refers to the statement not integrating the statements of the possible controlled subsidiaries or branches of the concerned bank with (without) a consolidated companion in Bankscope. C2 (C1) refers to the statement of a mother bank integrating the statements of its controlled subsidiaries or branches with (without) an unconsolidated companion in Bankscope.

private non-financial sector) from BIS. After plotting the concentration measures over time for each country, some extreme changes in concentration from one year to the next are easily spotted. Using the data on the shares of aggregate-level total assets and total credit, the number of banks in each year from Bankscope, and the ECB estimates on bank concentration for comparison, if the extreme changes in concentration in earlier years are caused by poor data coverage, then the country-year pairs are dropped. Table B.5 in Appendix B.4.3 shows the Bankscope data coverage (mean values for the shares of aggregates over time) after dropping country-year pairs with poor data coverage. The data descriptions for each EU or OECD country including the number of observations and the number of different types of banks are shown in Table B.4 in Appendix B.4.3.

Quarterly 5-year credit default swap (CDS) spreads for EU or OECD banks are from Thomson Reuters EIKON.[34] There are 218 unique banks in EU or OECD countries that have quarterly 5-year CDS spreads data available in the EIKON database. Each bank can have multiple CDS securities, with different seniorities, currencies, restructuring events, or data providers, which are uniquely identified in the database. Only one CDS security is kept for each bank. The cleaning procedures can be found in Appendix B.4.1. To analyse the relationship between banks' default probabilities proxied by the CDS spreads and their equity ratios, the cleaned CDS dataset is merged with the quarterly bank-level data on financial information from Bankscope.[35] The difficulty in merging the two datasets is that using the common identifiers (i.e., ISIN number and Ticker) can only allow me to match a limited number of banks since some banks are unlisted and some have missing ISIN or Ticker information in Bankscope. So for banks that cannot be matched by the identifiers, I manually match the banks from the two data sources using bank names. In this way, 174 banks can be matched, of which around 65% are commercial banks. I only keep 6 types of banks, i.e., bank holdings & holding companies, commercial banks, cooperative banks, finance companies, real estate & mortgage banks, and savings banks.[36] The period covered, number of observations in each quarter and other statistics for each country in the merged sample can be found in Table B.6 in Appendix B.4.3.

Annual country-level variables such as real GDP growth rate and inflation rate (growth rate of GDP deflator) are from World Bank. Quarterly real GDP growth rates are from OECD. Dollar/euro exchange rates used to convert the total assets of credit institutions into dollar values are obtained from the Federal Reserve Bank of St. Louis (FRED). Country-level lending rates used for model calibration are monetary and financial institution (MFI) interest rates from ECB. Table B.2 in Appendix B.4 summarizes the data sources used in this paper.

[34] Monthly spreads are also downloaded and averaged to give the quarterly spreads, which are very similar to the quarterly spreads and do not affect the results.

[35] Quarterly Bankscope data has a poor coverage as many banks do not report interim statements. However, this is not a problem if only looking at a small sample of large banks with CDS spreads data available.

[36] I drop 4 investment banks, 1 multi-lateral governmental banks, and 12 specialized governmental credit institutions. The final sample has 157 banks.

3.6 Empirical Evidence

The model predicts that when banks retain their profits to build up capital buffer, less banking competition improves financial stability measured by banks' default probabilities. I empirically assess this prediction in two steps. The first step is to test whether banking competition has an impact on the change in bank equity, where bank concentration is used as an inverse proxy for banking competition based on the Cournot model. The second step is to test if banks' equity ratios are negatively related to banks' default probabilities proxied by the CDS spreads. Since only a small sample of large banks have CDS data available, quarterly bank-level data on financial statements are used to allow for more data points. Following these two steps, two main empirical specifications based on the theoretical model are shown in Section 3.6.1 and 3.6.2. Finally, in Section 3.6.3, I also assess the model prediction in one step by investigating the direct relationship between banking competition and banks' default probabilities.

3.6.1 Imperfect Bank Competition and Change in Bank Equity

Following the dynamics of bank's equity accumulation (3.9), bank j's equity $n_{j,t}$ is the sum of the equity in the previous period $n_{j,t-1}$ and the realized net profit net of any dividends $D_{j,t}$ paid.[37] Equivalently, after rearranging (3.9),

$$\frac{n_{j,t} + D_{j,t} - R_{t-1}n_{j,t-1}}{k_{j,t-1}} = R_{b,t-1}G(\epsilon_t) - (R_{t-1} + \tau_j) \tag{3.29}$$

where $G(\epsilon_t) \equiv [1 - F(\bar{\omega}_t(\epsilon_t))] + \frac{1-\mu}{\bar{\omega}_t(\epsilon_t)} \int_0^{\bar{\omega}_t(\epsilon_t)} \omega f(\omega) d\omega < 1$ is the fraction of $R_{b,t}k_{j,t}$ earned by bank j when the aggregate shock takes a value of ϵ_t. The right hand side of the equation (3.29) is the profit margin that is negatively related to the number of banks N, since the equilibrium loan rate decreases with N (Proposition 1). As a result, a lower N or higher concentration raises the pre-dividend change in equity $\frac{n_{j,t}+D_{j,t}-n_{j,t-1}}{k_{j,t-1}}$ by raising the equilibrium loan rate and hence the profit margin. The observed equity $n_{j,t}$ from the bank's balance sheet is net of the cash dividend. As long as some positive net profits are retained as equity, then the change in equity $\frac{n_{j,t}-n_{j,t-1}}{k_{j,t-1}}$ is expected to be larger for a smaller N or less banking competition. Based on equation (3.29), the following empirical specification is used in the baseline analysis:

$$\frac{n_{j,c,t} - n_{j,c,t-1}}{k_{j,c,t-1}} = \beta_0 + \beta_1 N_{c,t-1} + \boldsymbol{\beta' X} + \beta_j + \beta_t + \beta_c + \varepsilon_{j,c,t} \tag{3.30}$$

[37] In this paper, dividends $D_{j,t}$ are paid via cash or share repurchase, in which case the dividend payment leads to a reduction in total equity. In reality, another way to pay dividend is through stock dividend (issuing more shares), which does not reduce total equity and simply results in a reallocation of equity funds, that is, retained earnings decrease and paid-in-capital increases by the same amount. The reason to use total equity as a proxy for $n_{j,t}$ instead of retained earnings is because total equity is a more relevant measure for capital buffer and in addition, around 53% of observations for EU countries would be lost if using retained earnings.

where j, c and t denote bank, country, and year respectively, and β_j, β_c, and β_t denote bank, country and year fixed effects respectively. $\boldsymbol{X}$ is a vector of bank-level and country-level control variables and $\boldsymbol{\beta'}$ is a row vector of the coefficients associated with each element in $\boldsymbol{X}$. In the baseline results shown in Table 3.2, the change in equity over lagged assets $\frac{n_{j,t}-n_{j,t-1}}{k_{j,t-1}}$ is used as the dependent variable. For robustness check, I also use $\frac{n_{j,t}+D_{j,t}-n_{j,t-1}}{k_{j,t-1}}$ as the dependent variable, where $D_{j,t}$ is proxied by cash dividends.

Since lagged bank concentration (proxy for $N_{c,t-1}$) is the main variable of interest that varies on country-year level, a pooled sample of different countries is used to control for the year fixed effects and exploit the cross-country variation.[38] Lagged number of banks $N_{c,t-1}$ is proxied by lagged Herfindahl Hirschman Index (HHI) or lagged 5-bank asset concentration ratio as one robustness check, as can be seen in Table B.8 in Appendix B.5. The vector $\boldsymbol{X}$ includes lagged loan impairment charge to gross loans ratio at the bank-year level, inflation rate (measured by the growth rate of GDP deflator) and lagged real GDP growth rate at the country-year level. Summary statistics of the key variables are shown in Table B.3 in Appendix B.4.3.

Table 3.2 shows the results by regressing the change in total equity over lagged total assets on lagged HHI, controlling for lagged loan impairment charge to gross loans ratio (loan impairment ratio), lagged real GDP growth rate and inflation rate (i.e., the growth rate of GDP deflator). The measure HHI (ECB) is directly obtained from the ECB, while HHI (Bankscope) is calculated from Bankscope annual data. Lagged loan impairment ratio and lagged GDP growth rate capture the variable $G(\epsilon_t)$ in equation (3.29) as they reflect the borrowers' ability to repay and hence the potential bank revenue loss due to non-performing loans. Controlling for inflation rate is because the dependent variable is not deflated. A higher inflation rate could inflate the change in equity in nominal terms and hence it should be positively related to the dependent variable. Banks from EU countries and OECD countries are used as two separate samples. Results without the controls are also shown in Table 3.2 for comparison.

It can be seen from Table 3.2 that bank concentration has a significant positive effect on the change in equity over lagged assets, as expected. Column 2 shows that when HHI (ECB) measure increases by 0.01 (or 10% from its mean of 0.1 across EU countries), the change in bank equity for EU banks increases by 0.0011 (or 14% relative to the mean change in bank equity of around 0.008 for EU banks). HHI calculated using Bankscope data gives smaller coefficients than the ECB measure, which can be explained by the differences in the data sources. Figure B.1 in Appendix B.4.3 compares the HHI from my own calculation with the HHI estimates from the ECB. As can be seen from Figure B.1, although the two measures have similar time variation in many EU countries such as Czech Republic, France, Greece, Italy, Latvia, Lithuania, and Spain, in general, the HHI from the ECB tends to be smaller in mangnitude than the one calculated using

[38]Regressions run separately for each country with year fixed effects will absorb the concentration variable.

Table 3.2: The Effect of Bank Concentration (HHI) on Change in Total Equity over Lagged Total Assets in EU and OECD Countries during 1999-2014

	(1) EU	(2) EU	(3) EU	(4) EU	(5) OECD	(6) OECD
L.HHI (ECB)	0.14***	0.11***				
	(0.02)	(0.02)				
L.HHI (Bankscope)			0.05***	0.04***	0.04***	0.03***
			(0.01)	(0.01)	(0.00)	(0.00)
L.loan impairment ratio		-0.06***		-0.07***		-0.15***
		(0.02)		(0.02)		(0.01)
L.GDP growth rate		0.11***		0.12***		0.06***
		(0.01)		(0.01)		(0.01)
inflation rate		0.11***		0.11***		0.12***
		(0.02)		(0.02)		(0.01)
Observations	44,419	44,419	45,033	45,033	199,317	199,317
Number of banks	4,875	4,875	4,936	4,936	19,230	19,230
Adjusted R^2	0.270	0.279	0.265	0.275	0.105	0.110
Within R^2	0.004	0.015	0.001	0.015	0.001	0.008
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Bank-level clustered standard errors in parentheses

Data sources: Bankscope annual data, ECB, World Bank

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

Note: The table shows the results from regressing the change in total equity over lagged total assets on lagged concentration index HHI, controlling for lagged loan impairment ratio (computed as loan impairment charge/gross loans), lagged real GDP growth rate (based on GDP in constant 2010 US dollar) and inflation rate (growth rate of GDP deflator). HHI (ECB) refers to the ECB estimate of HHI based on total assets of credit institutions in EU countries. HHI (Bankscope) is calculated using 6 types of banks (i.e., bank holding companies, commercial banks, cooperative banks, finance companies, real estate & mortgage banks, and savings banks) from Bankscope annual data.

Bankscope data, which is potentially due to a larger sample of credit institutions used by ECB.[39]

The signs of the other variables are also consistent with expectations. A higher loan impairment ratio lowers the equity due to loan losses, as a result, it is negatively related to the change in equity, as shown in Table 3.2. A higher GDP growth rate implies that more entrepreneurs would be able to repay their loans so it is positively related to change in equity. Inflation rate has a positive coefficient, as a higher inflation rate leads to a higher change in equity in nominal terms.

Results from four main robustness checks are shown in Appendix B.5. First, the regressions in Table 3.2 are re-run using 5-bank concentration ratio as an alternative inverse measure for $N_{c,t}$. As shown in Table B.8, 5-bank ratio from ECB still has a

[39] The comparison between 5-bank concentration ratio from ECB and the ratio calculated using Bankscope data is shown in Figure B.2 in Appendix B.4.3. Similar patterns are also observed for the 5-bank concentration ratio. In spite of a larger magnitude of the ratio calculated using Bankscope data, its time variation resembles that of the ECB measure in many EU countries.

Table 3.3: The Effect of Bank Equity Ratio on Bank CDS Spread in EU, Eurozone and OECD Countries during 2003-2016

	(1) EU	(2) EU	(3) Eurozone	(4) Eurozone	(5) OECD	(6) OECD
L.equity ratio	-0.34***	-0.25**	-0.32**	-0.23*	-0.33***	-0.33***
	(0.11)	(0.11)	(0.12)	(0.12)	(0.10)	(0.10)
L.loan impairment ratio		0.59***		0.65***		0.56***
		(0.15)		(0.17)		(0.12)
L.GDP growth rate		-0.74***		-1.00***		-0.43***
		(0.18)		(0.18)		(0.14)
Observations	1,344	1,340	998	994	3,008	2,871
Number of Banks	50	50	38	38	108	104
Adjusted R^2	0.723	0.752	0.727	0.763	0.690	0.719
Within R^2	0.060	0.159	0.056	0.180	0.093	0.175
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Bank-level clustered standard errors in parentheses

Data sources: Thomson Reuters EIKON, Bankscope quarterly data, OECD

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

Note: The table shows the results from regressing 5-year CDS spreads on banks' equity ratios, controlling for loan impairment charge to gross loans ratios, and real GDP growth rate. Bank, country and quarter fixed effects are included in all regressions. Quarterly data are used and all variables are in decimal places. Lagged explanatory variables are used. The sample consists of 6 types of banks (i.e., bank holding companies, commercial banks, cooperative banks, finance companies, real estate & mortgage banks, and savings banks).

highly significant positive effect on change in equity over lagged assets, while the measure calculated using Bankscope data is not significant in the sample of EU countries. Second, the samples of EU countries and OECD countries are further split into Eurozone countries, non-Eurozone EU countries, and non-EU OECD countries. As can be seen in Table B.9, HHI still has a siginificant positive coefficient, except for the Euro area countries when HHI is calculated using Bankscope data, which may be because the ECB measure is more reliable. Besides, the results tend to suggest that HHI has a larger impact on the change in equity for non-Eurozone EU countries than the Eurozone countries. Third, instead of using post-dividend change in equity over lagged assets as the dependent variable, cash dividends are added back. That is, change in equity plus cash dividends over lagged assets $\frac{n_{j,t}+D_{j,t}-n_{j,t-1}}{k_{j,t-1}}$ is used as the dependent variable. As expected, HHI has a slightly larger impact on the pre-dividend change in equity, as shown in Table B.10. Fourth, the sample over the period 1999-2014 is split into three different periods, 1999-2006, 2006-2014, and 2010-2014 for EU countries. Using the ECB measures for HHI and 5-bank concentration ratio, the results show that HHI is not significant during the pre-crisis period 1999-2006, as can be seen from Table B.11.

3.6.2 Bank Equity Ratio and Default Probability

According to Proposition 5, banks' default probabilities are negatively related to banks' equity ratios. Using the CDS spreads to proxy for banks' default probabilities, the following empirical specification is used:

$$\text{CDS Spread}_{j,c,t} = \beta_0 + \beta_1 \frac{n_{j,c,t-1}}{k_{j,c,t-1}} + \boldsymbol{\beta'}\boldsymbol{X} + \beta_j + \beta_c + \beta_t + \varepsilon_{j,c,t} \tag{3.31}$$

where j, c, t denote bank, country and quarter respectively. $\boldsymbol{X}$ is a vector of bank-level and country-level control variables and $\boldsymbol{\beta'}$ is a row vector of the coefficients associated with each element in $\boldsymbol{X}$. β_j, β_c, and β_t denote bank, country and quarter fixed effects respectively. The main variable of interest, $\frac{n_{j,c,t-1}}{k_{j,c,t-1}}$, is proxied by lagged bank's equity to total assets ratio. The vector $\boldsymbol{X}$ includes lagged loan impairment charge to gross loans ratio at the bank-quarter level, and lagged real GDP growth rate. The summary statistics of the CDS spreads and bank equity ratios for each country can be seen in Table B.7 in Appendix B.4.3.

The sample is divided into different groups of countries, i.e., EU, Eurozone and OECD countries. Using banks from different samples of countries, Table 3.3 shows that banks' equity ratios have a negative effect on their CDS spreads over the period of 2003-2016, controlling for lagged loan impairment charge to gross loans ratios and lagged real GDP growth rates.[40] More specifically, focusing on columns 2, 4 and 6 in Table 3.3, when bank equity ratios increase from 10% to 11%, their CDS spreads would be reduced by around 23 to 33 basis points, which represents around 10% to 15% of the mean CDS spread of around 220 basis points across EU banks. Table 3.3 also shows that a higher loan impairment charge to gross loans ratio leads to an increase in the CDS spread since it indicates a higher proportion of non-performing loans. A higher real GDP growth rate that implies a higher repay capacity of borrowers leads to a lower CDS spread, as shown in Table 3.3.

Robustness checks using different time periods (i.e., 2003-2011 and 2011-2016) and different data frequency (i.e., annual data) are shown in Table B.12 and B.13 in Appendix B.5. Table B.12 shows that equity ratios are not significant during 2003-2011 for EU countries and Eurozone countries. Using annual data instead of quarterly data does not change the results much if equity ratios are the only explanatory variable, as shown in Table B.13, however, it tends to reduce the magnitude and the significance of the coefficients on equity ratios after controlling for other variables. Another robustness check is to use country-year fixed effects instead of the quarter fixed effects to control for any country-level macroeconomic variables that vary over time and any potential time trend

[40] Hasan et al. (2016) find that market leverage (book value of liabilities over the sum of book value of liabilities and market value of equity) has a positive effect on banks' CDS spreads using a sample of 161 global banks during 2001-2011. Similarly, Acosta Smith et al. (2017) find that Tier 1 equity-to-total assets ratio has a negative effect on bank distress probabilities using data on a binary bank distress variable for EU banks during 2005-2014.

in the equity ratio variable. As shown in Table B.14 in Appendix B.5, the coefficients are significantly negative at 10% level, but the magnitude of the coefficients is smaller for EU and Eurozone countries.

3.6.3 Imperfect Bank Competition and Default Probability

In this section, I investigate whether imperfect banking competition lowers banks' default probabilities using a one-step approach. Table 3.4 shows the results from regressing banks' annual CDS spreads (proxy for banks' default probabilities) on bank concentration which is used as an inverse proxy for banking competition. In this section, annual CDS spreads (end of the fourth quarter data) are used since bank concentration has an annual frequency.

Table 3.4: Direct Relationship between CDS Spread and Concentration Measures in EU Countries

	(1) EU 2003-2016	(2) EU 2003-2011	(3) EU 2011-2016	(4) EU 2003-2016	(5) EU 2003-2011	(6) EU 2011-2016
L.HHI (ECB)	-0.08	-0.03	-0.52***			
	(0.06)	(0.09)	(0.11)			
L.equity ratio	-0.04	-0.33*	0.05	-0.05	-0.33*	0.02
	(0.05)	(0.19)	(0.08)	(0.05)	(0.19)	(0.08)
L.loan impairment ratio	0.50**	1.12***	0.24	0.50**	1.12***	0.22
	(0.21)	(0.36)	(0.15)	(0.22)	(0.36)	(0.15)
L.GDP growth rate	-0.08	-0.31**	-0.06***	-0.08	-0.31**	-0.08***
	(0.08)	(0.14)	(0.02)	(0.08)	(0.14)	(0.02)
L.5-bank ratio (ECB)				-0.03	-0.02	-0.16***
				(0.03)	(0.03)	(0.05)
Observations	702	342	422	702	342	422
Number of Banks	76	65	76	76	65	76
Adjusted R^2	0.683	0.605	0.866	0.684	0.606	0.863
Within R^2	0.093	0.245	0.226	0.095	0.246	0.211
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Bank-level clustered standard errors in parentheses

Data sources: Thomson Reuters EIKON, ECB, Bankscope annual data, World Bank

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

Note: The table shows the results from regressing 5-year CDS spreads on concentration index HHI or 5-bank concentration ratio, controlling for banks' equity ratios, loan impairment charge to gross loans ratios, and real GDP growth rate and including bank, country and year fixed effects. Annual data are used and all variables are in decimal places. Lagged explanatory variables are used. The sample consists of EU banks and is divided into different sub-samples based on time periods.

As can be seen from Table 3.4, the concentration index HHI or the 5-bank asset concentration ratio (both obtained from the ECB) has a significant negative effect on banks' CDS spreads during the post-crisis period (2011-2016).[41] More specifically, when

[41]The two different sample periods are divided by 2011 to make sure that the number of observations in each sample is roughly the same. The results are robust to dividing the whole sample by 2010 or 2012.

HHI (5-bank concentration ratio) increases by 0.01 or 10% (2%) from its mean of 0.1 (0.6) across EU banks, the CDS spreads would be reduced by around 52 (16) basis points or 24% (7%) from its mean of 220 basis points across EU banks during 2011-2016. Bank concentration is only significant during the post-crisis period because the cross-country variation in CDS spreads during the pre-crisis period is small. The finding is consistent with the model prediction that in the presence of bank equity accumulation, imperfect banking competition improves financial stability by lowering banks' default probabilities. The signs of the other explanatory variables align with the expectation, as discussed in Section 3.6.2.

The result is robust to using the Bankscope measures of concentration, as shown in Table B.15 in Appendix B.5. HHI still has a significant negative effect on banks' CDS spreads during the post-crisis period and 5-bank concentration ratio is significantly negative across all different sample periods. Results for OECD countries using the concentration measures from Bankscope are very similar to those shown in Table B.15. Finally, excluding banks' equity ratios gives very similar results, despite the positive correlation between the equity ratio and bank concentration.

3.7 Conclusions

This paper provides new theoretical and empirical evidence on the effects of imperfect competition in the banking sector on banks' equity ratios and thereby financial stability, which is measured through banks' default probabilities. By building a model of imperfect banking competition featuring bank equity accumulation, this paper finds that less banking competition can lead to a large gain in financial stability, provided that banks retain the greater profits as equity over time. As a result, macroprudential policies, for example, by limiting banks' dividend distribution to shareholders, can help ensure a larger gain in financial stability from less banking competition.

However, in the short run, a reduction in banking competition can jeopardize financial stability by lowering banks' equity ratios. For instance, by allowing solvent banks to merge with distressed banks to improve financial stability after a crisis, the merged banks have greater market power and hence more loan assets, resulting in lower equity-to-assets ratios and therefore higher default probabilities.

In addition, this paper quantifies the financial stability gain from less banking competition compared to the macroeconomic efficiency loss. In doing so, I find that bank equity accumulation is important for understanding the trade-off between financial stability and macroeconomic efficiency. In the absence of bank equity accumulation, i.e., when there is only the static margin effect, the gain in financial stability from less banking competition is very limited and is always outweighed by the macroeconomic efficiency loss. In this case, perfect banking competition is the best. However, when banks accumulate equity over time, the financial stability gain from less banking competition can be large enough to outweigh the macroeconomic efficiency loss, depending on the degree of banking competition.

More specifically, when there is very little competition, the macroeconomic efficiency loss is very large and completely outweighs any financial stability gain. For example, with a monopoly bank, the expected output is 40% lower compared to that with a perfectly competitive banking sector. Moving away from the extreme case (i.e., when there are more than six banks), the financial stability gain from imperfect banking competition can outweigh the macroeconomic efficiency loss.

Using data for EU and OECD countries during 1999-2016, I find supporting evidence for the model's prediction that when banks use retained earnings to build up their capital buffer, less banking competition improves financial stability measured through banks' default probabilities. I assess this prediction in two steps. First, bank concentration, an inverse measure for banking competition, has a significant positive effect on the change in bank equity. Second, banks' equity ratios have a negative effect on their default probabilities, which are proxied by the credit default swap spreads. Combining these two steps into one step, I find that bank concentration has a significant negative effect on banks' default probabilities during the post-crisis period, which is consistent with the model prediction.

As a result, this paper has shown from both a theoretical and empirical perspective the importance of imperfect banking competition on financial stability.

Chapter 4

Financial Frictions and Capital Misallocation

4.1 Introduction

Capital misallocation has important implications on aggregate productivity (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2008) and understanding the causes of capital misallocation is one of the central topics in the literature (Gopinath et al., 2017; Midrigan and Xu, 2014; Bartelsman et al., 2013).[1] Capital misallocation refers to the inefficient distribution of existing capital stock across producers given their productivity. As a contributing factor for capital misallocation, financial frictions or credit market imperfections have received a lot of attention.[2] More recently, there are a few papers aiming to quantify the impact of financial frictions on capital misallocation by estimating a structural model (Bai et al., 2018; David and Venkateswaran, 2017; Midrigan and Xu, 2014).

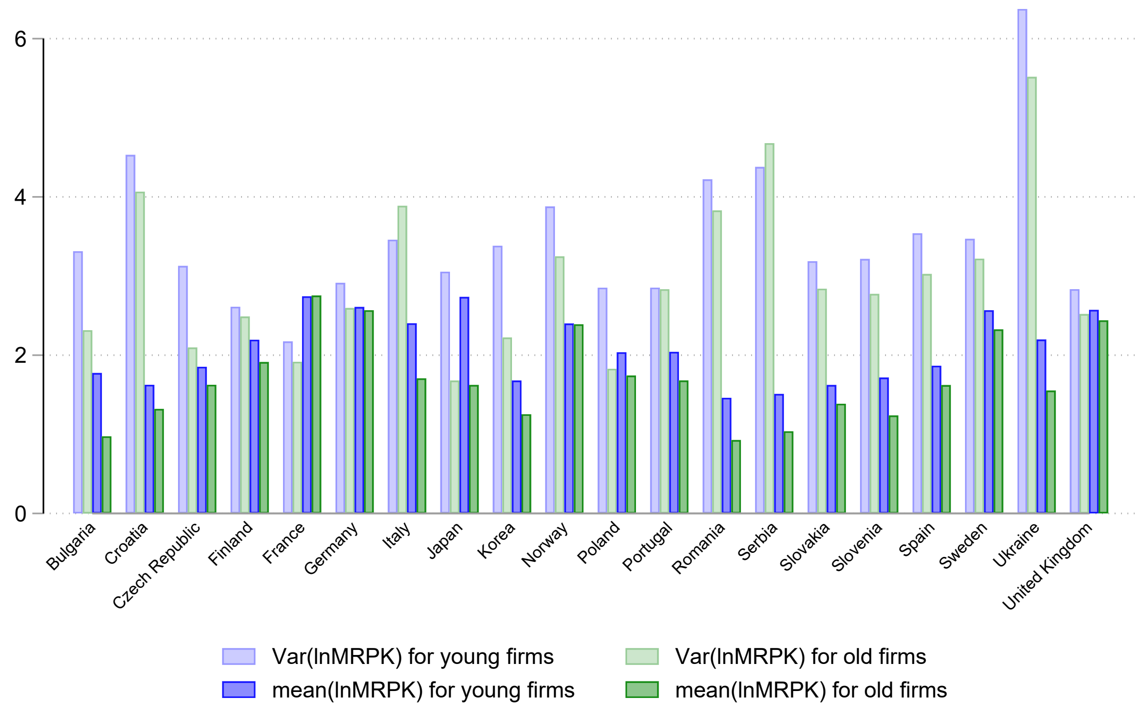
In this paper, I provide an alternative approach to quantify the impact of financial frictions on capital misallocation, which requires fewer restrictive assumptions and uses more information from large firm-level datasets. The approach consists of two steps. First, firms are classified as financially constrained or unconstrained using a switching regression approach. The idea is that the investment of the two types of firms follows two different processes, since the investment of constrained firms should be more responsive to cash flow than that of unconstrained firms. The probability of a firm being constrained is used to classify firms and is jointly estimated with the two different investment regimes.

Second, assuming that the distribution of the observed marginal revenue product of capital (MRPK) is a mixture of two distributions, one for each type of firm, this paper provides a statistical decomposition of capital misallocation, which is measured by the dispersion (variance) of MRPK. Since the efficient allocation of capital in a neoclassical

[1]See also Busso et al. (2013), Restuccia and Rogerson (2013).

[2]The potential impact of financial frictions on resource misallocation and total factor productivity losses has received a lot of attention in the recent literature. See Gopinath et al. (2017), Gilchrist et al. (2013), Banerjee and Duflo (2005), etc.

Figure 4.1: Dispersions and Means of Marginal Revenue Product of Capital (MRPK) in 2015



Note: The bar chart shows the cross-section variances (or dispersions) and means of the logarithm of the MRPK for young firms (age $\leqslant$ 15 years) and old firms (age $>$ 15 years) in each of the 20 selected countries in 2015. Firm age is computed as the difference between the year (2015) and the incorporation year plus one. MRPK (output elasticity of capital multiplied by output over capital stock based on a Cobb-Douglas production function) is computed as the nominal revenue (proxy for output) divided by fixed tangible assets (proxy for capital stock). Since the output elasticity of capital does not affect the dispersion of the MRPK within a given industry, it is neglected in the computation of the MRPK. Data source: Orbis

model indicates equalisation of the MRPK across firms, capital misallocation can be indirectly measured by the dispersion of the MRPK across firms within a given industry (Restuccia and Rogerson, 2017). The decomposition is motivated by the fact that younger firms who are more likely to be financially constrained have a higher dispersion and mean of MRPK than older firms, as can be seen in Figure 4.1. For most countries, it is highly statistically significant that both the means and the dispersions of MRPK for the young firms are larger than those for the old firms.[3]

Using the decomposition and the classified types estimated using large panels of manufacturing firms for 20 countries from the 1990s to 2015, this paper finds that the dispersions and means of MRPK for the financially constrained types are much larger

[3]One-sided t-tests and F-tests are used to test whether young firms have a higher mean and variance of log MRPK than old firms respectively in each country. The p-values from the t-test are smaller than 0.001 for all countries except for France, Germany, Norway, the UK, while the p-values from the F-test are smaller than 0.001 for all countries except for Croatia, Finland, Italy, Portugal, and Serbia. A small p-value for the t-test (F-test) rejects the null hypothesis that the means (variances) of ln MRPK are equal between the two types of firms.

than those for the unconstrained firms. For most countries and two-digit industries, more than a quarter of firms are classified as financially constrained and the presence of constrained firms can account for more than half of the observed dispersion of MRPK across firms.

This paper contributes to two strands of literature, namely, the macro literature on capital misallocation and the empirical finance literature on the impact of financial frictions on firm investment.

Financial frictions are often regarded as one of the leading contributing factors for capital misallocation and there are a few papers attempting to quantify the impact of financial frictions on capital misallocation by estimating a structural model. The importance of financial frictions is either implied by the estimated parameters or the predictions of quantitative models. However, whether financial frictions would cause large aggregate productivity loss via the capital misallocation channel remains unclear (Wu, 2018), which is likely due to the different modelling assumptions and datasets used.

There is evidence that financial frictions play an important role in generating the dispersion of MRPK in Spain and China. For instance, Gopinath et al. (2017) find that a size-dependent borrowing constraint is essential in generating the large increase in the dispersion of MRPK among Spanish manufacturing firms during 1999 and 2007. David and Venkateswaran (2017) use a quantitative model to find that firm-specific factors that are correlated with productivity, including financial frictions, account for around 47% of the MRPK dispersion using data for Chinese manufacturing plants from 1998 to 2009. Similar evidence can be found in Bai et al. (2018) who estimate their model using Chinese private manufacturing plants during 1998-2007. However, there are also papers showing that financial frictions only cause moderate efficiency losses through capital misallocation.[4] For instance, Midrigan and Xu (2014) calibrate their model using Korean manufacturing plants during 1991-1996 and find that financial frictions in the form of borrowing constraints do not lead to substantial aggregate productivity losses via resource misallocation.

My paper contributes to this strand of literature by proposing a new method to estimate the impact of financial frictions on capital misallocation, which relies on fewer restrictive assumptions and thus can be readily applied to a large number of countries. More specifically, I come up with a credit distortion measure using the decomposition of the dispersion of MRPK, which measures the fraction of the dispersion of MRPK that can be attributed to the presence of financially constrained firms. I then compute this credit distortion measure in each two-digit manufacturing industry from 20 countries during the period of the 1990s to 2015.

[4]Gilchrist et al. (2013) adopt a slightly different approach by applying an accounting framework (in which firm-specific borrowing costs are mapped into measures of resource misallocation) to US listed manufacturing firms and find that financial frictions are unlikely to be a major factor for resource misallocation, which is not very surprising as large listed firms tend to have better access to credit than small unlisted firms.

I build a simple model of firm dynamics with capital adjustment costs in the form of a one-period time to build for capital and a financial friction that gives rise to a borrowing constraint (due to costly debt enforcement as in Kiyotaki and Moore (1997)) to derive the empirical specifications in this paper. Based on the model, the investment of an unconstrained firm is driven by its expected future sales or productivity growth, while that of a constrained firm is driven by its cash flow. As a result, the MRPK for constrained and unconstrained firms is determined by different processes.

Given the distribution of the observed MRPK is a mixture of two distributions, one for each type of firms, I decompose the dispersion of MRPK across all firms into the dispersions and means of MRPK within the unconstrained and constrained groups of firms. This statistical decomposition gives new insights into the mechanisms through which the presence of financially constrained firms affects the extent of the observed dispersion of MRPK. While the usual mechanism emphasizes that the higher MRPK of constrained firms relative to the unconstrained firms would lead to a higher dispersion of MRPK, the decomposition in this paper shows that the dispersions of MRPK within different groups of firms also matter.[5] More importantly, this paper provides a new credit distortion measure using this decomposition, which measures the fraction of the dispersion of MRPK that is caused by the presence of constrained firms, but requires information on firms' financially constrained status.

The empirical finance literature on financial frictions and firm investment has proposed various ways to classify firms into constrained and unconstrained groups.[6] One common approach is to divide firms based on one indirect proxy for financial constraints, such as dividend payout, age, size or leverage (e.g., Moshiriana et al., 2017; Carpenter and Guariglia, 2008; Alti, 2003; Hubbard et al., 1995; Fazzari et al., 1988). As a direct extension of this approach, there are a lot of index-based measures of financial constraints that have been built on various combinations of firm characteristics (e.g., Mulier et al., 2016; Hadlock and Pierce, 2010; Whited and Wu, 2006; Lamont et al., 2001).[7]

Alternatively, instead of identifying constrained firms based on some a priori criteria, a switching regression model could be used to simultaneously estimate the probability that

[5]For example, the mechanism in Gopinath et al. (2017) operates through the increasing gap in the MRPK between the constrained and unconstrained firms. In response to a decline in the real interest rate, unconstrained firms increase their capital demand and experience a decline in their MRPK, while the constrained firms are not able to invest more capital and their MRPK does not fall, leading to an increased dispersion of the MRPK in the sector.

[6]To test whether financial frictions affect firm investment, classification of firms into constrained and unconstrained groups is often a prerequisite step after which the differential investment behavior between the two groups of firms can be tested.

[7]Lamont et al. (2001) construct the KZ index, which is a weighted sum of five accounting ratios, using the regression coefficients from Kaplan and Zingales (1997) as the weights. An alternative index measuring the degree of financial constraints (WW index) was constructed by Whited and Wu (2006), based on estimating an investment Euler equation from a structural model. Hadlock and Pierce (2010) question the validity of the KZ index and WW index and propose a new measure based on firm size and age only, arguing that these two firm characteristics are particularly useful in predicting the levels of financial constraints.

firms are financially constrained and the two different investment regimes for constrained and unconstrained firms (e.g., Almeida and Campello, 2007; Hovakimian and Titman, 2006; Hu and Schiantarelli, 1998). The two investment regimes differ in terms of the sensitivity to cash flow. That is, firm investment should be more sensitive to cash flow for constrained firms after controlling for the investment opportunity. The probability estimated using this maximum likelihood approach can be used to classify firms into constrained and unconstrained groups.

In fact, the switching regression model is closely related to the index-based approach. The index-based measure of financially constrained status can also give the probability of a firm being constrained via a logit or probit function. However, this probability does not use any model structure or data information of the two investment regimes, unlike in a switching regression. This paper builds on the switching regression model to classify firms into constrained and unconstrained firms.

This paper contributes to this strand of the empirical finance literature by providing evidence for unlisted firms and more countries, using new proxies for investment opportunity that are motivated by the theoretical model. Existing literature often uses Tobin's q as a proxy for investment opportunity in a sample of US listed firms.[8] However, q is not available for unlisted firms, so this paper uses sales growth, value added growth and productivity growth as different proxies for investment opportunity, in order to analyse unlisted firms that are more likely to suffer from financial constraints. Furthermore, instead of focusing on the US firms, this paper provides new evidence for manufacturing firms in 20 countries from the 1990s to 2015.

The remainder of the paper is structured as follows. Section 4.2 shows the theoretical framework and the decomposition of the dispersion of MRPK. Section 4.3 describes the data and the summary statistics. Section 4.4 shows two different empirical specifications and the corresponding empirical results. Section 4.5 concludes.

4.2 Theoretical Framework

This section builds a simple model of firm dynamics with one-period time to build for capital and a borrowing constraint. The model is used to show that the capital demand and hence the marginal revenue product of capital (MRPK) for unconstrained and constrained firms are driven by different processes. I then decompose the dispersion of MRPK across all firms into the dispersions and means within the two types of firms. The model is also used to derive the two different investment equations for empirical analysis in Section 4.4.

[8] In a model with convex capital adjustment costs, marginal q or the shadow value of one additional unit of capital is a sufficient statistic for investment. However, as pointed out by Schiantarelli (1995), stock markets may be inefficient and if stock prices are driven by fads, q may not be a good proxy for investment opportunity.

4.2.1 Modeling Preliminaries

Assume there are M monopolistically competitive firms in a specific subsector s of the manufacturing industry, which are infinitely lived, each producing a differentiated product. Firms are indexed by i, where $i = 1, ..., M$.[9] For notational simplicity, sector subscripts are suppressed in this theory section. The total industry output y_t is a constant elasticity of substitution (CES) aggregate of M differentiated products:

$$y_t = \left(\sum_{i=1}^{M} y_{i,t}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{4.1}$$

where $y_{i,t}$ is the real output produced by firm i in period t, and $\epsilon > 1$ is the elasticity of substitution between varieties. Each firm i in period t produces output $y_{i,t}$ using capital $k_{i,t-1}$, which is predetermined (i.e., purchased and installed in period $t-1$), materials $m_{i,t}$, and labor $l_{i,t}$ via an industry-specific Cobb-Douglas production function:

$$y_{i,t} = A_{i,t} k_{i,t-1}^{\alpha_k} m_{i,t}^{\alpha_m} l_{i,t}^{\alpha_l} \tag{4.2}$$

where $A_{i,t}$ is the firm-specific physical productivity or total factor productivity (TFP), and $\alpha_k \in (0,1)$, $\alpha_m \in (0,1)$ and $\alpha_l \in (0,1)$ are the industry-specific output elasticities of capital, materials and labor, respectively. Assume constant returns to scale such that $\alpha_k + \alpha_m + \alpha_l = 1$.

Firms engage in monopolistic competition and each of them charges a price $p_{i,t}$ for their differentiated product i. Given the aggregate output index y_t (4.1), it can be calculated from the cost minimization problem of the buyers of the industry output that each firm faces a downward-sloping demand with a constant elasticity $\epsilon > 1$ for their product:

$$y_{i,t} = \left(\frac{p_{i,t}}{p_t} \right)^{-\epsilon} y_t \tag{4.3}$$

where both the industry output y_t and the industry price p_t are normalized to one in this partial equilibrium model, following Gopinath et al. (2017). As a result, combining the production function (4.2) and the demand for the firm's product (4.3), the revenue-based production function is:

$$p_{i,t} y_{i,t} = Z_{i,t} k_{i,t-1}^{\beta_k} m_{i,t}^{\beta_m} l_{i,t}^{\beta_l} \tag{4.4}$$

where $Z_{i,t} \equiv A_{i,t}^{\frac{\epsilon-1}{\epsilon}}$ is the revenue-based productivity or TFPR, and $\beta_k \equiv \alpha_k \frac{\epsilon-1}{\epsilon}$, $\beta_m \equiv \alpha_m \frac{\epsilon-1}{\epsilon}$ and $\beta_l \equiv \alpha_l \frac{\epsilon-1}{\epsilon}$ are the industry-specific revenue elasticities of capital, materials and labor respectively.[10] The revenue-based production function is often used in the

[9]This partial equilibrium model can be used to describe firm dynamics within the manufacturing industry, as well as any subsector of it, as used in the empirical analysis in Section 4.4.

[10]Note that $\beta_k + \beta_m + \beta_l = \frac{\epsilon-1}{\epsilon}$.

literature because firm-level prices $p_{i,t}$ and output $y_{i,t}$ are often unavailable, while $p_{i,t}y_{i,t}$ can be empirically measured by nominal revenue or sales.[11]

Assume that the firm-specific revenue-based productivity can be decomposed into the product of three independent components, so $Z_{i,t} \equiv Z_t z_i z_{i,t}$, with a common trend Z_t, a firm-specific component z_i and an idiosyncratic component $z_{i,t}$, where the latter follows an AR(1) process in logs:

$$\ln z_{i,t} = \rho \ln z_{i,t-1} + e_{i,t} \tag{4.5}$$

with $\rho \in (0,1)$ indicating the persistence of the process, and $e_{i,t} \sim N(0, \sigma_z^2)$ being an independent and identically normally distributed random variable with mean zero and variance σ_z^2.

4.2.2 Firm's Capital Choice and Financial Frictions

Assume firms own the capital, which depreciates at a rate $\delta \in [0,1]$. They also purchase and install new capital each period for production in the following period. Assuming they start with different levels of initial net worth $n_{i,0}$ at $t = 0$, firms with low initial net worth may need to borrow at an exogenous real gross interest rate R_0 to finance the purchase of physical capital $k_{i,0}$. Similarly, firms with enough net worth to finance the capital save at the same interest rate.[12] Firms install the purchased capital $k_{i,0}$, and at the beginning of $t = 1$, the productivity shocks realize and output $y_{i,1}$ is produced using the installed capital $k_{i,0}$, labor $l_{i,1}$ and materials $m_{i,1}$. Assume firms hire labor and acquire materials in a competitive market at an exogenous real wage rate w_t and real price of materials $p_{m,t}$ in each period t. Let $n_{i,t}$ denote the firm's net worth before its choice of $k_{i,t}$ and any borrowing $b_{i,t}$ in period t.

Financial friction is modelled via a costly debt enforcement problem, based on Kiyotaki and Moore (1997). In other words, borrowers cannot be forced to repay unsecured debt. Since creditors recognize the possibility of default by borrowers, they would never lend more than what they can obtain in the case of default. Hence, each firm would face a borrowing constraint that is tied to the value of their collateral, which is the value of their undepreciated capital:

$$b_{i,t} \leqslant \phi(1-\delta)k_{i,t} \tag{4.6}$$

where $\phi \in (0,1)$ is the loan-to-value ratio.[13] Assume firms are risk-neutral and in each period t, after the production of output and the payments of wage, materials and debt, there is a constant probability $\varphi \in (0,1)$ that the firm exits, in which case the firm

[11] The notion of TFPR was introduced by Foster et al. (2008). The heterogeneity in TFPR across firms can reflect a combination of productivity differences and monopolistic pricing distortions.

[12] This paper abstracts from other financial frictions such as imperfect banking competition, as analysed in Chapter 2 and 3.

[13] Since $b_{i,t}$ is used to buy part of the capital stock $k_{i,t}$, $k_{i,t}$ is not observed at the time of borrowing and there is a possibility that the firm absconds with the borrowed fund. To avoid this possibility, assume that the loan is conditional on the firm using it to purchase capital.

consumes its net worth $n_{i,t}$. The surviving firms choose how much capital to purchase in period t given their net worth $n_{i,t}$. The exiting firms are replaced by new firms, with random levels of initial net worth, such that the total number of firms in the industry stays unchanged.[14] Given the firm faces a borrowing constraint, it is reasonable to assume that the firm delays consumption until the period it exits. Let $\eta \in (0,1)$ denote the firm's discount factor. Each firm i in period t chooses its capital $k_{i,t}$, labor $l_{i,t}$ and materials $m_{i,t}$ to maximize the expected discounted terminal net worth:[15]

$$\mathrm{E}_t \sum_{\tau=0}^{\infty} \varphi(1-\varphi)^{\tau}\eta^{\tau} n_{i,t+1+\tau} \tag{4.7}$$

subject to the borrowing constraint (4.6). The net worth $n_{i,t+1}$ at the beginning of period $t+1$ equals the sum of the revenue $p_{i,t+1}y_{i,t+1}$ and the undepreciated capital stock $(1-\delta)k_{i,t}$, net of the real wage cost $w_{t+1}l_{i,t+1}$, the real materials cost $p_{m,t+1}m_{t+1}$, and the gross debt interest payment $R_t b_{i,t}$:

$$n_{i,t+1} \equiv p_{i,t+1}y_{i,t+1} - w_{t+1}l_{i,t+1} - p_{m,t+1}m_{i,t+1} - R_t b_{i,t} + (1-\delta)k_{i,t} \tag{4.8}$$

Using the assumption that firms do not consume until the period they exit and (4.8), it can be shown that the firm finances the purchase of capital using either the internal financing (net worth) or external financing (debt):[16]

$$k_{i,t} = n_{i,t} + b_{i,t} \tag{4.9}$$

Using (4.9) to rewrite the borrowing constraint (4.6) in terms of net worth:

$$k_{i,t} \leqslant \frac{n_{i,t}}{1-\phi(1-\delta)} \tag{4.10}$$

Let $\lambda_{i,t}$ denote the Lagrange multiplier associated with the borrowing constraint (4.10), and let $k^U_{i,t}$ and $k^C_{i,t}$ denote firm i's unconstrained capital demand and constrained capital demand, respectively. It is shown in Appendix C.1.3 and C.1.5 that the firm's capital demand $k_{i,t}$ is:

$$k_{i,t} = \begin{cases} k^U_{i,t} & \text{if} \quad \lambda_{i,t} = 0 \\ k^C_{i,t} & \text{if} \quad \lambda_{i,t} > 0 \end{cases} \tag{4.11}$$

[14] One example to justify this assumption is that in each period, a random fraction of households start new firms using their savings as initial net worth.

[15] In period 0, the firm only chooses capital $k_{i,0}$ because there is no capital yet to produce.

[16] Suppose firm i consumes $c_{i,t+1}$ in period $t+1$, then the firm faces the following budget constraint:

$$c_{i,t+1} + k_{i,t+1} = b_{i,t+1} + p_{i,t+1}y_{i,t+1} - w_{t+1}l_{i,t+1} - p_{m,t+1}m_{t+1} - R_t b_{i,t} + (1-\delta)k_{i,t} \equiv b_{i,t+1} + n_{i,t+1}$$

Given that the firm does not consume until the period of exit, $c_{i,t+1} = 0$ and hence $k_{i,t+1} = b_{i,t+1} + n_{i,t+1}$. In the terminal period T, $b_{i,T} = k_{i,T} = 0$, so $c_{i,T} = n_{i,T}$.

where the log of the capital demand of a constrained firm i is:

$$\ln k_{i,t}^{C} = \ln n_{i,t} - \ln[1 - \phi(1-\delta)] \tag{4.12}$$

and the log of the capital demand of an unconstrained firm i is:

$$\begin{aligned}\ln k_{i,t}^{U} =& \epsilon\rho \ln z_{i,t} + (1+\epsilon\beta_k)\left\{\ln\left(\beta_k \beta_l^{\frac{\epsilon\beta_l}{1+\epsilon\beta_k}} \beta_m^{\frac{\epsilon\beta_m}{1+\epsilon\beta_k}}\right) - \ln(r_t+\delta)\right. \\ &\left. + \ln \mathrm{E}_t\left[\left(\frac{Z_{t+1}}{w_{t+1}^{\beta_l} p_{m,t+1}^{\beta_m}}\right)^{\frac{\epsilon}{1+\epsilon\beta_k}}\right] + \frac{\epsilon}{1+\epsilon\beta_k}\ln z_i + \frac{\sigma_z^2\epsilon^2}{2(1+\epsilon\beta_k)^2}\right\}\end{aligned} \tag{4.13}$$

where $r_t \equiv R_t - 1$ is the net real interest rate. Since productivity is assumed to follow an AR(1) process, the expected future productivity can be written in terms of the current productivity $z_{i,t}$. As can be seen, the capital demand of the unconstrained firm is increasing in its idiosyncratic transitory productivity $z_{i,t}$, permanent productivity z_i, and the expected future common productivity Z_{t+1}, and decreasing in the net real interest rate r_t and the expected factor prices (w_{t+1} and $p_{m,t+1}$). Intuitively, higher (expected) firm productivity (both common, permanent and idiosyncratic) leads to a higher demand for physical capital. A higher net interest rate increases the marginal cost of capital and thus reduces the capital demand. A higher real wage or price of materials reduces the demand for labor and materials respectively, leading to a lower marginal revenue product of capital (MRPK). Hence, capital demand falls to ensure that the expected MRPK equals the user cost of capital ($r_t + \delta$). By contrast, the constrained firm cannot operate at an optimal scale and its capital demand is constrained by its net worth $n_{i,t}$, as shown in (4.12).

4.2.3 Dispersion in Marginal Revenue Product of Capital

Define firm i's period-t marginal revenue product of capital $\mathrm{MRPK}_{i,t}$ as:

$$\mathrm{MRPK}_{i,t} \equiv \frac{\partial p_{i,t} y_{i,t}}{\partial k_{i,t-1}} = \beta_k Z_{i,t} k_{i,t-1}^{\beta_k - 1} l_{i,t}^{\beta_l} m_{i,t}^{\beta_m} = \beta_k \frac{p_{i,t} y_{i,t}}{k_{i,t-1}} \tag{4.14}$$

where $p_{i,t} y_{i,t}$ denotes the nominal revenue. Based on the model, if all firms were financially unconstrained, their expected MRPK is identical, as they face the same interest rate. There is still dispersion in ex post MRPK due to the different realizations of the productivity shocks across firms, but this source of dispersion is not treated as misallocation in the literature, since the allocation of capital is efficient ex ante (e.g., Restuccia and Rogerson, 2017; Asker et al., 2014). Hence, capital misallocation should be measured by the dispersion of the expected MRPK.

However, it is difficult to measure the expected MRPK, so this paper measures capital misallocation by the static dispersion of MRPK across firms within a given two-digit industry.[17] As a result, capital misallocation may be overestimated if production inputs (such as capital in this paper) are chosen before the shock realizes. However, this paper does not attempt to disentangle the efficient causes of the dispersion. Assuming capital adjustment costs affect constrained and unconstrained firms equally, this paper only aims to estimate the proportion of the dispersion caused by the financial friction.

The static dispersion of MRPK across all firms, as shown in Appendix C.1.2, can be written as:

$$\begin{aligned}\text{Var}_i(\ln\text{MRPK}_{i,t}) =& \psi_1\text{Var}_i(\ln z_i) + \psi_1\text{Var}_i(e_{i,t}) + \psi_1\rho^2\text{Var}_i(\ln z_{i,t-1}) + \psi_2\text{Var}_i(\ln k_{i,t-1}) \\ & - \psi_3\text{Cov}_i(\ln z_i + \rho\ln z_{i,t-1}, \ln k_{i,t-1})\end{aligned} \tag{4.15}$$

where Var_i and Cov_i denote the cross-section variance and covariance across firms in a given time period, and $\psi_1 \equiv \left(\frac{\epsilon}{1+\epsilon\beta_k}\right)^2$, $\psi_2 \equiv \left(\frac{1}{1+\epsilon\beta_k}\right)^2$, and $\psi_3 \equiv 2\frac{\epsilon}{(1+\epsilon\beta_k)^2}$ are positive coefficients. This is a general decomposition that holds regardless of the types of firms. In general, capital misallocation measured by the static dispersion of MRPK depends on the cross-section dispersions of idiosyncratic permanent productivity $\text{Var}_i(\ln z_i)$, the productivity innovation $\text{Var}_i(e_{i,t})$, the past productivity $\text{Var}_i(\ln z_{i,t-1})$, and installed capital stock $\text{Var}_i(\ln k_{i,t-1})$, and the cross-section covariance between the firm's capital and different components of the firm's productivity, as shown in (4.15). There is no dispersion of the marginal revenue product of labor (MRPL) in this model, as $\text{MRPL}_{i,t} = \beta_l\frac{p_{i,t}y_{i,t}}{l_{i,t}} = w_t$, using the first order condition with respect to labor from the model.

In the absence of the one-period time to build and financial frictions, it can be shown that there is no dispersion of MRPK in this model.[18] In other words, there are two causes for the dispersion of MRPK: time-to-build for capital and financial frictions. First, due to a one-period time to build, $k_{i,t}$ is chosen based on the expected future productivity $\text{E}_tZ_{i,t+1}$. Hence, any realized productivity $Z_{i,t+1}$ that differs from the expectation would cause the MRPK to differ across firms ex post.[19] This explains why the cross-section dispersion in the productivity innovation $\text{Var}_i(e_{i,t})$ causes dispersion of MRPK. In fact, if all firms are unconstrained (i.e., without the financial frictions), $\text{Var}_i(e_{i,t})$ is the only source of dispersion of MRPK. Let $\text{Var}_i(\ln\text{MRPK}^U_{i,t})$ denote the cross-section variance of the log MRPK across financially unconstrained (U) firms, then it is shown in Appendix

[17]Note that $k_{i,t-1}$ is used in the model to reflect that capital is chosen in period $t-1$, but only used in period t. In empirical analysis, fixed tangible asset in period t is used to measure $k_{i,t-1}$, so the dispersion of MRPK (4.14) is still the static dispersion of MRPK.

[18]If firms are unconstrained and capital adjusts immediately in response to productivity shocks, firms can always borrow to finance their optimal demand for capital and their MRPK will be equalised within a given industry, as they face the same interest rate: $\text{MRPK}_{i,t} = (r_t + \delta)$.

[19]Restuccia and Rogerson (2017) note that one problem with measuring misallocation using the dispersion of marginal products is that when inputs are chosen before firm-specific shocks realize, the marginal products across firms may not equalize in every time period even under efficient allocation. Similarly, Asker et al. (2014) pointed out that in the presence of capital adjustment costs, the ex ante efficient choice of capital can be inefficient ex post.

C.1.4 that:

$$\text{Var}_i(\ln\text{MRPK}^U_{i,t}) = \psi_1 \text{Var}_i(e_{i,t}) \tag{4.16}$$

where the superscript U denotes unconstrained firms. As can be seen, the dispersion of MRPK for unconstrained firms is purely driven by the cross-section dispersion in the productivity innovation $\text{Var}_i(e_{i,t})$.

Second, financial frictions in the form of a collateral constraint also cause MRPK to differ across firms. Using the first order condition for capital, financially constrained firms have higher expected MRPK than unconstrained firms as they cannot borrow enough to finance their optimal capital demand, as shown in Appendix C.1. As a result, the differences of MRPK between constrained and unconstrained firms contribute to the overall dispersion of MRPK. Furthermore, this paper finds that the dispersion of MRPK within constrained firms is also important for understanding the overall dispersion of MRPK caused by financial frictions. Let $\text{Var}_i(\ln\text{MRPK}^C_{i,t})$ denote the cross-section variance of the log MRPK across financially constrained (C) firms, then it is shown in Appendix C.1.6 that:

$$\begin{aligned}\text{Var}_i(\ln\text{MRPK}^C_{i,t}) = &\psi_1 \text{Var}_i(\ln z_i) + \psi_1 \text{Var}_i(e_{i,t}) + \psi_1 \rho^2 \text{Var}_i(\ln z_{i,t-1}) + \psi_2 \text{Var}_i(\ln n_{i,t-1}) \\ &- \psi_3 \text{Cov}_i(\ln z_i + \rho \ln z_{i,t-1}, \ln n_{i,t-1})\end{aligned} \tag{4.17}$$

where the superscript C denotes constrained firms. Note that $\text{Var}_i(\ln\text{MRPK}^C_{i,t})$ now also depends on the variances and covariances of the logs of z_i, $z_{i,t-1}$ and $n_{i,t-1}$. Using (4.16) and (4.17), it can be shown that constrained firms have a higher dispersion of MRPK than unconstrained firms.[20] As can be seen from (4.17), the dispersion of MRPK for constrained firms is lower when firms' net worth and productivity are more positively correlated. The firm's productivity and net worth are expected to be positively correlated since more productive firms tend to be more profitable, and thus accumulate more net worth over time. Intuitively, firms with higher productivity (in period $t-1$ or permanently) anticipate higher future productivity and hence would want to demand more capital. Since more productive firms also tend to have more net worth, their borrowing capacity is higher, which means they are able to borrow more while being constrained, and their capital demand is closer to their optimal unconstrained capital demand, bringing down their MRPK and thereby also the dispersion of MRPK among constrained firms.

This paper analyzes the impact of financial frictions on the dispersion of MRPK by empirically estimating the percentage of the dispersion that can be attributed to the presence of constrained firms. Suppose there are N_t unconstrained firms in an industry

[20] Rearranging (4.17) gives:

$$\text{Var}_i(\ln\text{MRPK}^C_{i,t}) = \psi_1 \text{Var}_i(e_{i,t}) + \text{Var}_i(\psi_1^{\frac{1}{2}} \ln z_i + \psi_1^{\frac{1}{2}} \rho \ln z_{i,t-1} - \psi_2^{\frac{1}{2}} \ln n_{i,t-1})$$

where the second term on the RHS, which is strictly positive, is the only difference from the dispersion of MRPK for unconstrained firms (4.16). Hence, $\text{Var}_i(\ln\text{MRPK}^C_{i,t}) > \text{Var}_i(\ln\text{MRPK}^U_{i,t})$.

in a given time period t, and the remaining $M_t - N_t$ firms are constrained, then it is shown in Appendix C.2 that the cross-section variance of MRPK across all firms can be decomposed as follows:

$$\begin{aligned}\mathrm{Var}_i(\ln\mathrm{MRPK}_{i,t}) =& \frac{N_t}{M_t}\mathrm{Var}_i(\ln\mathrm{MRPK}^U_{i,t}) + \frac{M_t - N_t}{M_t}\mathrm{Var}_i(\ln\mathrm{MRPK}^C_{i,t}) \\ &+ \frac{N_t(M_t - N_t)}{M_t^2}\left[\mathrm{E}_i(\ln\mathrm{MRPK}^U_{i,t}) - \mathrm{E}_i(\ln\mathrm{MRPK}^C_{i,t})\right]^2\end{aligned} \tag{4.18}$$

where the cross-section variances and means on the right hand side of (4.18) are defined for the two subgroups of firms. For example, $\mathrm{E}_i(\ln\mathrm{MRPK}^U_{i,t})$ denotes the cross-section mean of MRPK across unconstrained firms only. As can be seen from (4.18), the overall dispersion of MRPK equals a weighted average of the dispersion for unconstrained and constrained firms plus a measure of distance between the mean for each group. It is shown in (4.16) and (4.17) that the dispersion of MRPK within the unconstrained group is driven by the dispersion in productivity innovation while that within the constrained group is also driven by the dispersions in firms' net worth $n_{i,t-1}$, firms' idiosyncratic permanent productivity z_i, the realized idiosyncratic transitory productivity $z_{i,t-1}$, and the covariance between their productivity and net worth.

Constrained firms have a higher MRPK than unconstrained firms because of the lower level of capital that can be financed.[21] As a result, the cross-section average of MRPK for constrained firms $\mathrm{E}_i(\ln\mathrm{MRPK}^C_{i,t})$ is larger than that for the unconstrained firms $\mathrm{E}_i(\ln\mathrm{MRPK}^U_{i,t})$.[22] According to the decomposition of the variance of MRPK (4.18), the larger the gap in mean MRPK between the two groups of firms $\left[\mathrm{E}_i(\ln\mathrm{MRPK}^U_{i,t}) - \mathrm{E}_i(\ln\mathrm{MRPK}^C_{i,t})\right]$, the higher the cross-section variance of MRPK at an industry-year level. This resembles the usual mechanism that the presence of borrowing constraints increases the dispersion of MRPK due to the differences in MRPK between unconstrained and constrained firms.

More importantly, based on the decomposition (4.18), if financially constrained firms can be identified empirically, then the proportion of the dispersion of MRPK that is caused by the financial friction can be estimated by:

$$\text{Credit Distortion} \equiv \frac{\mathrm{Var}_i(\ln\mathrm{MRPK}_{i,t}) - \frac{N_t}{M_t}\mathrm{Var}_i(\ln\mathrm{MRPK}^U_{i,t})}{\mathrm{Var}_i(\ln\mathrm{MRPK}_{i,t})} \in [0,1] \tag{4.19}$$

where $\mathrm{Var}_i(\ln\mathrm{MRPK}_{i,t})$ is observable in data, while $\mathrm{Var}_i(\ln\mathrm{MRPK}^U_{i,t})$ is the cross-section variance defined over the unconstrained group of firms, which can only be calculated once

[21]This follows from the first order condition for capital as shown in Appendix C.1.

[22]As shown in (C.12) in Appendix C.1.1, on top of the effect of lower capital for constrained firms on increasing their MRPK, the realizations of their productivity $Z_{i,t}$ are also different for constrained and unconstrained firms. More specifically, when the realized $Z_{i,t}$ is sufficiently high, a firm is more likely to be constrained as the firm demands more capital. As a result, the cross-section average productivity for constrained firms is higher than that for unconstrained firms, and this further explains why $\mathrm{E}_i(\ln\mathrm{MRPK}^C_{i,t}) > \mathrm{E}_i(\ln\mathrm{MRPK}^U_{i,t})$.

the unconstrained firms are identified. Section 4.4.2 uses a switching regression approach to identify the two groups of firms, following Hu and Schiantarelli (1998).

If the constrained firms were not present, so $N_t = M_t$ and $\text{Var}_i(\text{lnMRPK}_{i,t})$ equals $\text{Var}_i(\text{lnMRPK}^U_{i,t})$ (as implied by (4.19)), then the credit distortion measure equals zero. If all firms were constrained, so $N_t = 0$, then the credit distortion measure equals one. Hence, the fraction of the dispersion of MRPK caused by the presence of constrained firms is based on the difference between the observed $\text{Var}_i(\text{lnMRPK}_{i,t})$ and the counterfactual variance of MRPK without constrained firms, $\text{Var}_i(\text{lnMRPK}^U_{i,t})$, which is normalised by the overall dispersion in MRPK.[23]

The credit distortion measure (4.19) may also capture any other structural differences across the two types of firms that lead to different dispersions of MRPK within the unconstrained and constrained firms. For instance, although this paper assumes a constant markup (i.e., $\frac{\epsilon}{\epsilon-1} > 1$) for all firms, in practice, larger firms tend to have higher markups, while smaller firms tend to have lower markups as they are more likely to operate in an environment that is close to perfect competition. If the markup dispersion is higher for larger, unconstrained firms, then $\text{Var}_i(\text{lnMRPK}^U_{i,t})$ is also higher, leading to a lower credit distortion measure. The econometric analysis in Section 4.4.2 therefore aims to control for other structural differences between the two types of firms, to mitigate this issue.

Furthermore, this paper only considers a quantity-based credit distortion measure and does not incorporate any distortions in the price of credit such as the interest rate wedge between the borrowing and saving rate caused by imperfect banking competition.[24] Thus, the current measure of credit distortion is likely to underestimate the full impact of credit distortions by neglecting that constrained firms also tend to face greater price distortions.

In addition, this paper assumes that the revenue elasticity of capital β_k is the same for all firms within a subsector of the manufacturing industry. As a result, β_k does not affect the dispersion of MRPK and is neglected when computing the MRPK. However, if β_k differs across unconstrained and constrained firms, this also contributes to the dispersion of MRPK for the two types of firms and thus affecting the credit distortion measure. Using (4.14) and the first order condition for materials, MRPK can be equally measured using nominal revenue $p_{i,t}y_{i,t}$ or value added $\text{VA}_{i,t}$:

$$\text{MRPK}_{i,t} = \beta_k \frac{p_{i,t}y_{i,t}}{k_{i,t-1}} = \frac{\beta_k}{1-\beta_m}\frac{\text{VA}_{i,t}}{k_{i,t-1}} \tag{4.20}$$

where $\text{VA}_{i,t} \equiv p_{i,t}y_{i,t} - p_{m,t}m_{i,t} = (1-\beta_m)p_{i,t}y_{i,t}$. Using nominal value added over fixed tangible assets to measure MRPK requires both the revenue elasticities of materials β_m and capital β_k to be identical across firms, which is more restrictive. As a result, this

[23]Note that $\text{Var}_i(\text{lnMRPK}^U_{i,t})$ is multiplied by the fraction of unconstrained firms $\frac{N_t}{M_t}$ to ensure that the measure is always positive.

[24]The effect of imperfect banking competition on the interest rate wedge when borrowers are financially constrained is analysed in Chapter 2.

paper measures MRPK using nominal revenue over fixed tangible assets in the baseline analysis and uses nominal value added to measure MRPK for robustness checks.

4.3 Data

The firm-level data for different countries used in this paper are from the Bureau van Dijk's Orbis historical financial database, which provides annual financial information from firms' balance sheets and income statements from early 1990s to 2015.[25] The financial variables extracted from the Orbis historical financial database are combined with some time-invariant variables extracted from the Orbis rolling 10 years database. After extracting the variables from these two databases, the dataset for each country is cleaned following similar procedures as outlined in Kalemli-Ozcan et al. (2015). The full cleaning procedure and the summary statistics for the sample used for empirical analysis are shown in Appendix C.4.

One major advantage of this database is that it contains both listed and unlisted firms, unlike Compustat or Worldscope which only cover listed firms. Since unlisted firms are more likely to suffer from financial constraints than listed firms which tend to be larger, it is useful to study financial constraints using datasets that contain those unlisted firms. The empirical analysis in this paper focuses on the manufacturing sector in each country because the capital stock can be proxied by fixed tangible assets, whereas in other industries, it is more difficult to define the capital stock.[26] Within the manufacturing sector, an industry is defined by either a two-digit or four-digit NACE Rev.2 code in this paper.

Table 4.1 shows some basic information on the datasets of the manufacturing sector for each of the 20 countries used in the baseline analysis, including the time period covered, total number of observations, average number of firms per year, unique number of four-digit industries, and the fraction of observations coming from unlisted firms. Since value added is used to calculate the marginal revenue product of capital (MRPK) for robustness checks and value added is often less available than sales, the countries in the baseline sample are selected based on the total number of observations and the availability of value added.[27] The statistics reported in Table 4.1 are after dropping the observations with missing operating revenue and missing or zero fixed tangible assets,

[25]The time series for some European countries start in 1990, while for many countries, the time series are shorter.

[26]Based on the two-digit NACE Rev.2 code, the manufacturing sector is in the range of 10 to 33. The descriptions for each two-digit industry can be found in Table C.5 in Appendix C.4.

[27]Countries are ranked according to their total number of observations and also their availability of value added (in terms of the percentage of the total observations). The two ranks have equal weight and the top 19 countries according to the weighted rank are selected plus Japan. Japan has low availability for value added but it has a large number of observations and it is an important country to look at since this paper focuses on the manufacturing firms. US is not selected due to the low availability of value added and the relatively low number of observations.

Table 4.1: Data Description for Each Country in the Baseline Sample

Country	Period	Observations	Obs/Year	Industries	Unlisted Firms
Bulgaria	1995-2015	119,346	5,683	223	0.983
Croatia	1998-2015	124,184	6,899	220	0.981
Czech Republic	1994-2015	176,420	8,019	289	0.995
Finland	1995-2015	163,600	7,790	227	0.992
France	1995-2015	1,316,144	62,674	229	0.994
Germany	1990-2015	255,056	9,810	298	0.975
Italy	1995-2015	1,716,653	81,745	302	0.998
Japan	1989-2015	593,512	21,982	199	0.959
Korea	2001-2015	817,068	54,471	198	0.973
Norway	1996-2015	109,826	5,491	217	0.990
Poland	1994-2015	167,273	7,603	236	0.981
Portugal	1998-2015	372,214	20,679	227	0.999
Romania	1995-2015	558,739	26,607	231	0.984
Serbia	1999-2015	165,237	9,720	235	0.930
Slovakia	1995-2015	76,190	3,628	228	0.980
Slovenia	1997-2015	93,570	4,925	213	0.991
Spain	1994-2015	1,428,899	64,950	230	0.999
Sweden	1997-2015	299,408	15,758	229	0.988
Ukraine	2001-2015	422,144	28,143	227	0.985
United Kingdom	1994-2015	294,092	13,368	230	0.966

Note: The sample from each country consists of manufacturing firms only. Period shows the time period covered in each cleaned country-specific dataset. Observations and Obs/year show the total number of firm-year observations and the average number of firms respectively during the period covered in a given country. Industries shows the number of unique four-digit NACE Rev.2 industries over the period covered in each country. The last column shows the fraction of observations coming from unlisted firms.

but before dropping those with missing value added. As can be seen from Table 4.1, in most countries, more than 98% of the observations are from unlisted firms.

Table 4.2 shows the medians of some main variables of interest in this paper. As can be seen, the median number of employees is below 20 in 16 out of 20 countries, which shows that the dataset contains a lot of small firms. The year-on-year change in the log of fixed tangible assets FTA (proxy for capital stock) measures the firm net investment, which is used as the dependent variable in the empirical analysis shown in Section 4.4. The median value of ΔlnFTA is negative as small firms account for a large proportion of the sample. The average ΔlnFTA is positive in most countries, as can be seen in Table C.2 in Appendix C.4.

According to the model, the year-on-year change in the log of sales, the log of value added, or the log of productivity can be used as different proxies for the investment opportunity, as shown in (4.21) and (4.22) in Section 4.4.1. Nominal value added is computed as the difference between operating revenue and material costs for most

Table 4.2: Medians of Selected Variables for Each Country in the Baseline Sample

Country	Employees	Age	ΔlnFTA	ΔlnSales	ΔlnVA	ΔlnTFPR	Var(logMRPK)
Bulgaria	14	9	-0.026	0.047	0.048	-0.002	2.86
Croatia	5	12	-0.038	0.020	0.031	-0.007	3.70
Czech Republic	23	12	-0.026	0.034	0.035	-0.001	2.80
Finland	8	15	-0.066	0.026	0.033	0.007	2.18
France	6	14	-0.102	0.024	0.026	0.004	1.76
Germany	27	21	-0.033	0.012	0.038	0.009	2.97
Italy	12	15	-0.033	0.020	0.026	0.001	3.23
Japan	17	32	-0.024	0.009	0.018		2.07
Korea	17	9	-0.000	0.084	0.101	0.005	2.79
Norway	9	12	-0.059	0.039	0.043	0.011	3.46
Poland	90	13	-0.019	0.048	0.052	0.003	2.34
Portugal	7	14	-0.074	0.012	0.019	0.001	2.47
Romania	6	9	-0.003	0.108	0.137	0.003	2.98
Serbia	5	11	-0.008	0.095	0.130	-0.008	3.94
Slovakia	15	10	-0.053	0.027	0.033	-0.001	2.92
Slovenia	5	15	-0.060	0.034	0.034	0.001	2.69
Spain	8	13	-0.038	0.027	0.034	-0.000	2.52
Sweden	5	17	-0.077	0.031	0.031	0.007	2.78
Ukraine	11	9	-0.040	0.076	0.109	-0.009	5.21
United Kingdom	76	17	-0.039	0.030	0.050	0.005	2.55

Note: The sample from each country consists of manufacturing firms only. All the statistics reported in the table except for the last column are medians. Employees shows the number of employees for each country. Age is computed as the difference between year and incorporation year plus one. ΔlnFTA, ΔlnSales, ΔlnVA, and ΔlnTFPR denote the year-on-year change in the log fixed tangible assets, log sales, log value added and log productivity respectively. TFPR is estimated using the Wooldridge (2009) approach. TFPR cannot be estimated for Japan due to the lack of data on material costs. Var(lnMRPK) is the cross-section variance of the log marginal revenue product of capital, where MRPK is computed as nominal revenue over fixed tangible assets. The variance reported here is unconditional on industries.

countries.[28] If firm productivity were estimated accurately, then this would be a more exogenous measure than sales growth. However, there is no perfect measure for firm productivity. This paper uses productivity growth as a robustness check, which is estimated using the Wooldridge (2009) approach.[29]

The last column of Table 4.2 shows the dispersion in log marginal revenue product of capital MRPK. According to (4.14), $\text{MRPK}_{i,t}$ is calculated as nominal revenue $p_{i,t}y_{i,t}$ over the capital stock $k_{i,t-1}$.[30] It can be seen that there is a large variation in the dispersion

[28]The original 'value added' variable in Orbis is the sum of taxation, profit/loss for the period (equivalent to profit/loss after taxation plus the extraordinary and other profit/loss), cost of employees, depreciation and interest paid. The computed value added is used if it has more observations. Except for Germany, Japan (no data for materials costs), Portugal, Spain, and UK, the computed value added is used in the other countries in the baseline sample.

[29]The details of this estimation approach can be found in Appendix C.3.

[30]According to (4.14), $\text{MRPK}_{i,t}$ also depends on the revenue elasticity of capital β_k. However, since β_k is often estimated at a two-digit industry level to ensure enough number of observations and hence is

of MRPK across countries. Ukraine has the highest dispersion of MRPK, followed by Serbia and Croatia, whereas France has the lowest dispersion of MRPK.

Table 4.3: Medians of Selected Variables for Listed and Unlisted Firms

	Employees		Age		ΔlnFTA		ΔlnSales	
Country	Unlisted	Listed	Unlisted	Listed	Unlisted	Listed	Unlisted	Listed
Bulgaria	13	201	8	46	-0.026	-0.026	0.047	0.027
Croatia	5	261	11	56	-0.039	-0.020	0.020	0.014
Czech Republic	23	750	12	14	-0.026	-0.024	0.034	0.026
Finland	8	957	15	26	-0.067	-0.004	0.025	0.042
France	6	342	14	26	-0.103	0.009	0.024	0.049
Germany	26	752	21	44	-0.035	0.005	0.012	0.039
Italy	12	413	15	27	-0.033	-0.003	0.020	0.034
Japan	15	784	31	61	-0.025	-0.009	0.008	0.025
Korea	15	168	8	21	-0.000	0.014	0.085	0.067
Norway	9	406	12	14	-0.059	0.000	0.039	0.068
Poland	87	225	13	20	-0.019	0.022	0.047	0.074
Portugal	7	336	14	41	-0.074	-0.033	0.012	0.018
Romania	5	231	8	14	-0.003	-0.004	0.108	0.083
Serbia	4	111	10	20	-0.009	-0.000	0.098	0.067
Slovakia	15	225	10	14	-0.054	-0.034	0.027	0.008
Slovenia	5	228	15	23	-0.060	-0.011	0.034	0.028
Spain	8	558	13	40	-0.039	0.002	0.027	0.046
Sweden	5	81	17	18	-0.078	-0.010	0.030	0.068
Ukraine	10	328	9	21	-0.041	-0.009	0.075	0.111
United Kingdom	73	470	16	21	-0.040	0.008	0.029	0.045

Note: The sample from each country consists of manufacturing firms only. The table shows the median values of four variables computed using subsamples of unlisted and listed firms. Employees shows the median number of employees for each country. Age is computed as the difference between year and incorporation year plus one. ΔlnFTA and ΔlnSales denote the year-on-year change in the log fixed tangible assets and log sales respectively.

Table 4.3 shows the median values computed using subsamples of unlisted and listed firms. As can be seen, the median number of employees and age for listed firms are much larger than for unlisted firms. The year-on-year change in the log of fixed tangible assets ΔlnFTA is also higher for listed firms except for Bulgaria and Romania. By contrast, there is no clear pattern for the year-on-year change in the log of sales (i.e., sale growth) between the listed and unlisted firms. In 8 out of 20 countries, the sales growth of unlisted firms is higher than that of listed firms.

the same across firms within the industry, it does not matter for the dispersion of log MRPK within a two-digit industry. As discussed in Section 4.4.1, when using value added to compute MRPK, β_m is assumed to be the same for each subsector as well, so it is better to use sales revenue to compute MRPK.

4.4 Empirical Analysis

Starting from Fazzari et al. (1988), there is a large literature on testing whether firm investment responds to cash flow fluctuations after controlling for the investment opportunity proxied by Tobin's q. By dividing firms into different groups according to some firm characteristics that can affect their constrained status, a higher investment-cash flow sensitivity within the 'constrained' group relative to the 'unconstrained' group after controlling for q would suggest the presence of financial frictions. Since the datasets contain mostly unlisted firms for which q is not available, this paper uses sales growth, value added growth and productivity growth as different proxies for the investment opportunity. Section 4.4.1 shows that these proxies are motivated by the model in Section 4.2.

Using ex ante divisions of firms into constrained and unconstrained groups based on the marginal revenue product of capital (MRPK), Section 4.4.1 shows some evidence that the investment of constrained firms is more sensitive to their internal financing. Section 4.4.2 uses a switching regression model to classify firms into constrained and unconstrained firms, where the probability of a firm being constrained depends on multiple firm characteristics and is estimated jointly with two different investment regimes depending on the firm's constrained status.

4.4.1 Firm Investment and Financial Frictions

Following (4.13), the capital demand of an unconstrained firm is determined by its productivity and the factor prices (r_t, w_{t+1} and $p_{m,t+1}$). Let $\Delta \ln k^U_{i,t} \equiv \ln k^U_{i,t} - \ln k^U_{i,t-1}$ denote the net capital investment of an unconstrained firm.[31] It can be shown that:

$$\Delta \ln k^U_{i,t} = \epsilon\rho \Delta \ln z_{i,t} - (1+\epsilon\beta_k)\Delta \ln(r_t+\delta) + (1+\epsilon\beta_k)\Delta \ln \mathrm{E}_t\left[\left(\frac{Z_{t+1}}{w^{\beta_l}_{t+1} p^{\beta_m}_{m,t+1}}\right)^{\frac{\epsilon}{1+\epsilon\beta_k}}\right] \tag{4.21}$$

where $\Delta \ln z_{i,t}$ is the idiosyncratic revenue-based total factor productivity growth that can proxy for the firm's investment opportunity. Note that $k_{i,t-1}$ is the capital decided in period $t-1$ but used in period t, so it is measured by contemporaneous fixed tangible assets $\mathrm{FTA}_{i,t}$ in the data. As a result, (4.21) implies that $\Delta \ln \mathrm{FTA}_{i,t+1}$ depends positively on productivity growth $\Delta \ln z_{i,t}$ and change in log expected future common productivity, and negatively on change in log user cost of capital $\Delta \ln(r_t+\delta)$ and change in log expected future real wage and real price of materials. The last two terms in (4.21) are common to all firms in each subsector, so they are absorbed by (four-digit) industry-year fixed effects.

[31] Note that from the model, gross investment is $k_{i,t} - (1-\delta)k_{i,t-1}$. Since $\Delta \ln k_{i,t} = \ln \frac{k_{i,t}}{k_{i,t-1}} = \ln\left(\frac{k_{i,t}-k_{i,t-1}}{k_{i,t-1}} + 1\right) \approx \frac{k_{i,t}-k_{i,t-1}}{k_{i,t-1}}$, $\Delta \ln k_{i,t}$ is a measure for growth rate of capital stock or net investment normalised by the capital stock $k_{i,t-1}$.

The investment of unconstrained firms can also be written in terms of the expected growth of the nominal revenue $p_{i,t}y_{i,t}$. Using the first order condition for capital, the investment of an unconstrained firm is:

$$\Delta \ln k_{i,t}^{U} = \Delta \ln \mathrm{E}_t \left[p_{i,t+1}y_{i,t+1}\right] + \ln \frac{r_{t-1}+\delta}{r_t+\delta} \tag{4.22}$$

Since $\Delta \ln \mathrm{E}_t \left[p_{i,t+1}y_{i,t+1}\right] = \Delta \ln \mathrm{E}_t \left[(1-\beta_m)p_{i,t+1}y_{i,t+1}\right]$, both the sales growth and the value added growth can be used to proxy for $\Delta \ln \mathrm{E}_t \left[p_{i,t+1}y_{i,t+1}\right]$. If the firm's productivity can be estimated accurately, then it is a more exogenous measure for investment opportunity. However, given the existing empirical methods may not perfectly back out the unobserved firm's revenue-based total factor productivity, I also use lagged sales growth and value added growth to proxy for the investment opportunity in the regression analysis.[32] Since expected future sales or value added growth is unavailable in data and current growth is likely to cause endogeneity problems, lagged growth is used instead under the assumption that lagged growth can predict future growth.

Following (4.12), the capital demand of a constrained firm is only determined by its net worth and hence its investment is determined by the growth in its net worth.

$$\Delta \ln k_{i,t}^{C} = \Delta \ln n_{i,t} \tag{4.23}$$

Alternatively, it can be expressed in terms of cash flow $\mathrm{CF}_{i,t}$, which is defined as the revenue net of the wage payments, material costs and net interest payments on debt, $\mathrm{CF}_{i,t} \equiv p_{i,t}y_{i,t} - w_t l_{i,t} - p_{m,t}m_{i,t} - r_{t-1}b_{i,t-1}$, assuming that all debt is rolled over in each period with no repayment of principal until the terminal period. It is shown in Appendix C.1.5 that:

$$\Delta \ln k_{i,t}^{C} \approx \frac{k_{i,t}^{C} - k_{i,t-1}^{C}}{k_{i,t-1}^{C}} = \frac{1}{1-\phi(1-\delta)} \frac{\mathrm{CF}_{i,t}}{k_{i,t-1}} - \frac{\delta}{1-\phi(1-\delta)} \tag{4.24}$$

where the firm's cash flow is the sum of its net income (equivalent to the change in net worth $\Delta n_{i,t}$ in this model setup) and depreciation of the capital stock. Since the capital $k_{i,t-1}$ decided in period $t-1$ but used in period t is measured by contemporaneous fixed tangible assets $\mathrm{FTA}_{i,t}$ in the data, (4.24) implies that $\Delta \ln \mathrm{FTA}_{i,t+1}$ depends positively on $\frac{\mathrm{CF}_{i,t}}{\mathrm{FTA}_{i,t}}$. Regressing $\Delta \ln \mathrm{FTA}_{i,t}$ on $\frac{\mathrm{CF}_{i,t-1}}{\mathrm{FTA}_{i,t-1}}$ would cause simultaneity bias as $\mathrm{FTA}_{i,t-1}$ appears on both sides of the equation. Hence, I use lagged cash flow over twice lagged fixed tangible assets $\frac{\mathrm{CF}_{i,t-1}}{\mathrm{FTA}_{i,t-2}}$ in empirical analysis to prevent the simultaneity bias. This paper uses cash flow to measure the availability of internal financing because it is likely to be a more exogenous measure than net worth as capital stock forms part of the net

[32] There is no need to deflate the nominal sales or nominal value added because four-digit industry*year dummies are included in the regressions, which will absorb the deflators varying at the two-digit industry-year level (as deflators are often only available at the two-digit industry level in the data).

worth. As can be seen from (4.24), whenever the firm is constrained by external financing, its investment is completely determined by the availability of internal financing.

According to the model, in the absence of any financial frictions, the investment in capital should not be affected by the availability of internal financing, after controlling for the investment opportunity. By contrast, when financial frictions exist so that some firms are financially constrained due to low net worth, these firms' investment is expected to be sensitive to the availability of their internal financing.

However, there are two caveats to the above implication. First, a firm's constrained status can change over time. If a firm was constrained in the previous period but becomes unconstrained this period, then the firm's investment will still depend on the net worth even after controlling for the investment opportunity.[33] More importantly, since there is no perfect empirical measure for the investment opportunity, the availability of internal financing also captures current profitability and is thus related to expected future profitability (due to persistence in $z_{i,t}$). Hence, when running a regression with the change in fixed tangible asset on investment opportunity and cash flow, it is likely to get a significant effect on cash flow even for unconstrained firms. Nevertheless, if the investment-cash flow sensitivity decreases with the factors that alleviate financial frictions, then this sensitivity is likely to be linked to financial frictions (Ağca and Mozumdar, 2008).

To test this hypothesis, I use the model to identify firms that are likely to be constrained. As explained in Section 4.2.3, with the financial frictions, a firm with a higher MRPK is more likely to be constrained due to a lower level of capital. An indicator variable d is created to indicate whether the firm belongs to the constrained group. I then run the following regression for each country separately:

$$\begin{aligned}\Delta\ln\mathrm{FTA}_{i,s,t} =& \gamma_0 + \gamma_1\mathrm{Opp}_{i,s,t-1} + \gamma_2\mathrm{Opp}_{i,s,t-1} * d_{i,s,t} + \gamma_3\frac{\mathrm{CF}_{i,s,t-1}}{\mathrm{FTA}_{i,s,t-2}} + \gamma_4\frac{\mathrm{CF}_{i,s,t-1}}{\mathrm{FTA}_{i,s,t-2}} * d_{i,s,t} \\ &+ d_{i,s,t} + \gamma_i + \gamma_{s,t} + \varepsilon_{i,s,t}\end{aligned} \tag{4.25}$$

where i, s, and t denote firm, (four-digit) industry and year, respectively, and $\mathrm{Opp}_{i,s,t-1}$ denotes the investment opportunity, which is proxied by lagged sales growth, value added growth or productivity growth. The availability of internal financing $\frac{\mathrm{CF}_{i,s,t-1}}{\mathrm{FTA}_{i,s,t-2}}$ is measured by lagged cash flow over twice lagged fixed tangible assets $\mathrm{FTA}_{i,s,t-2}$, where cash flow is the sum of profit for the period and depreciation. γ_i and $\gamma_{s,t}$ represent firm and four-digit industry*year fixed effects respectively.

Firms with higher MRPK are more likely to be financially constrained and hence their investment should be more dependent on the availability of their internal financing and thus more sensitive to cash flow, as shown in (4.24). As a result, γ_4 is expected to be significantly positive if $d_{i,s,t}$ indicates these firms. Similarly, γ_2 is expected to be

[33] Note that in this case, $\Delta\ln k_{i,t} \equiv \ln k^U_{i,t} - \ln k^C_{i,t}$, so the firm's investment depends on $\ln z_{i,t}$ and $\ln n_{i,t-1}$.

significantly negative as the investment of the 'constrained' firms should be less affected by investment opportunity than their unconstrained counterparts. However, lagged sales or value added growth are noisy proxies for investment opportunity, so the results on investment opportunity may not be reliable.

Table 4.4 shows the results from regressing firm investment $\Delta \text{lnFTA}_{i,t}$ on lagged sales growth $\Delta \text{logSales}_{i,t-1}$ and lagged cash flow $\text{CF}_{i,t-1}$ over twice lagged fixed tangible assets $\text{FTA}_{i,t-2}$ for each country separately, where both explanatory variables are interacted with an indicator variable that equals one if the firm's lagged log of MRPK is in the top 30% and zero if otherwise, following the specification (4.25). A higher MRPK tends to indicate a more constrained status. The effect of cash flow on investment is expected to be larger for firms with higher MRPK. The effect of investment opportunity would only be smaller for these firms if there were a perfect measure for investment opportunity.

As can be seen from Table 4.4, in all countries except for Japan, the coefficient for cash flow interacted with the indicator variable for MRPK is highly significant and positive, indicating that the investment of firms with higher MRPK is more sensitive to their cash flow, as expected. In 9 out of 20 countries, these firms' investment also responds less to the investment opportunity, as shown by the significantly negative coefficient on lagged sales growth interacted with the indicator variable. However, in Ukraine, the coefficient for the interaction term between lagged sales growth and the indicator variable is significantly positive, which is inconsistent with the hypothesis. Besides, in three countries, the coefficient on lagged sales growth is significantly negative. These anomalous findings on lagged sales growth are likely because lagged sales growth is a noisy proxy for investment opportunity.

To interpret the coefficients in terms of their economic significance, I use France as an example. More specifically, as shown in Table 4.4, the coefficient of cash flow is 0.017 for France, which means for older firms, when cash flow increases by 0.1 (or 6.3% from its unconditional mean of 1.6), the capital growth increases by 0.0017 (or 8.1% from its unconditional mean of -0.021). The coefficient of the cash flow interacted with the age dummy of 0.006 means that for firms with MRPK in the top 30%, when cash flow increases by 0.1, their investment increases by 0.0023 or 11% from its unconditional mean. Similarly, the investment of older firms increases by 0.000074 or 0.4% from its unconditional mean of -0.021 when the lagged sales growth rate increases by 0.001 (or 3.6% from its unconditional mean of 0.028). By contrast, for the same increase in the lagged sales growth, the investment of firms with a higher MRPK increases by only 0.000035 or around 0.17% from its unconditional mean. The summary statistics of the key variables can be found in Table C.2 and C.3 in Appendix C.4.

For robustness checks, I use different proxies for investment opportunity. Results using lagged value added growth are very similar to the baseline results, so they are not shown in the Appendix due to space limitations. Using lagged productivity growth gives more insignificant or anomalous findings on the coefficient for the interaction term

Table 4.4: Capital Investment-Cash Flow Sensitivity and Marginal Revenue Product of Capital (MRPK)

Country	ΔlnSales	ΔlnSales $* d$	$\frac{CF}{FTA}$	$\frac{CF}{FTA} * d$	$d(\text{MRPK} > p70)$	Within R^2	Observations
Bulgaria	0.001	-0.014	-0.001	0.022***	0.400***	0.0668	67,519
	(0.0045)	(0.0131)	(0.0019)	(0.0024)	(0.0118)		
Croatia	0.034***	0.010	0.002	0.022***	0.414***	0.0687	82,909
	(0.0047)	(0.0144)	(0.0016)	(0.0020)	(0.0103)		
Czech Republic	0.012**	-0.050***	0.004***	0.019***	0.386***	0.0657	119,733
	(0.0048)	(0.0142)	(0.0015)	(0.0019)	(0.0081)		
Finland	0.018***	-0.018*	0.007***	0.006***	0.282***	0.0535	118,999
	(0.0041)	(0.0108)	(0.0013)	(0.0015)	(0.0060)		
France	0.074***	-0.039***	0.017***	0.006***	0.306***	0.0693	1,011,014
	(0.0030)	(0.0076)	(0.0004)	(0.0005)	(0.0021)		
Germany	0.026***	-0.008	0.001	0.008***	0.289***	0.0516	68,887
	(0.0087)	(0.0187)	(0.0014)	(0.0017)	(0.0094)		
Italy	0.035***	0.006	0.011***	0.012***	0.295***	0.0514	1,282,096
	(0.0016)	(0.0039)	(0.0006)	(0.0007)	(0.0020)		
Japan	-0.007	-0.053***	0.042***	-0.007	0.235***	0.0391	59,519
	(0.0071)	(0.0189)	(0.0099)	(0.0110)	(0.0098)		
Korea	0.018***	-0.060***	0.004***	0.028***	0.588***	0.1045	392,415
	(0.0025)	(0.0073)	(0.0009)	(0.0011)	(0.0057)		
Norway	0.037***	-0.017	0.006***	0.010***	0.363***	0.0624	81,944
	(0.0064)	(0.0200)	(0.0011)	(0.0013)	(0.0092)		
Poland	0.018***	-0.027*	0.008***	0.008***	0.284***	0.0505	87,532
	(0.0059)	(0.0150)	(0.0020)	(0.0023)	(0.0081)		
Portugal	0.039***	0.008	0.005***	0.014***	0.305***	0.0454	273,761
	(0.0030)	(0.0092)	(0.0012)	(0.0014)	(0.0044)		
Romania	-0.006***	-0.009*	0.005***	0.014***	0.396***	0.0541	367,514
	(0.0020)	(0.0049)	(0.0006)	(0.0008)	(0.0046)		
Serbia	0.023***	-0.003	0.005***	0.019***	0.384***	0.0663	105,557
	(0.0032)	(0.0091)	(0.0015)	(0.0020)	(0.0086)		
Slovakia	-0.019**	-0.035*	0.005	0.026***	0.457***	0.0614	47,523
	(0.0077)	(0.0189)	(0.0032)	(0.0041)	(0.0143)		
Slovenia	-0.118***	0.014	0.001	0.031***	0.422***	0.0513	62,121
	(0.0104)	(0.0161)	(0.0028)	(0.0036)	(0.0140)		
Spain	0.021***	0.002	0.012***	0.017***	0.276***	0.0521	1,111,449
	(0.0016)	(0.0050)	(0.0007)	(0.0008)	(0.0021)		
Sweden	0.044***	-0.001	0.010***	0.005***	0.304***	0.0488	237,819
	(0.0038)	(0.0110)	(0.0007)	(0.0009)	(0.0051)		
Ukraine	0.017***	0.009**	-0.001	0.008***	0.313***	0.0343	286,337
	(0.0017)	(0.0047)	(0.0008)	(0.0009)	(0.0053)		
United Kingdom	0.032***	-0.030***	0.003***	0.004***	0.266***	0.0432	200,281
	(0.0047)	(0.0106)	(0.0007)	(0.0008)	(0.0051)		

Note: The table shows the coefficients from regressing $\Delta\text{lnFTA}_{i,t}$ on lagged sales growth $\Delta\text{lnSales}_{i,t-1}$ and lagged cash flow over twice lagged fixed tangible assets $\frac{\text{CF}_{i,t-1}}{\text{FTA}_{i,t-2}}$, and each of which interacted with a dummy that equals one if lagged log MRPK is in the top 30% and zero if otherwise. The last column shows the number of firm-year observations used in each regression. Firm and four-digit industry*year fixed effects are included in all regressions. Firm-level clustered standard errors are reported in parentheses.

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

between the lagged productivity growth and the indicator variable, but the coefficient on the interaction term between cash flow and the indicator is still highly significant and positive in all countries except for the UK, as shown in C.6 in Appendix C.6.

Instead of using MRPK, other variables could be used to try to identify constrained firms, such as size or age.[34] However, using $d_{i,s,t}$ to indicate either small or young firms yields inconclusive results.[35] This is likely because firm size and age are imperfectly related to MRPK which is the theoretical determinant for the firm's constrained status.

4.4.2 Switching Regression Model

One problem with the specification (4.25) in Section 4.4.1 is that classifying firms as constrained or unconstrained based on just one variable may not be sufficient. As a result, more recent papers studying the investment-cash flow sensitivities tend to use switching regression models of Maddala (1986) and jointly estimate the probability of a firm being constrained and two different investment regimes depending on whether the firm is constrained or not (e.g., Almeida and Campello, 2007; Hovakimian and Titman, 2006; Hu and Schiantarelli, 1998). The objective of these papers is to show that financial frictions matter for firm investment, avoiding the use of ex ante classification of firms based on one firm characteristic each time.

The main objective of this paper is to estimate the percentage of the dispersion of MRPK that is caused by financial frictions or the presence of constrained firms based on the decomposition (4.18), which requires classifying firms as constrained or unconstrained. Consequently, this paper uses the switching regression model not only to show the importance of financial frictions in affecting firm investment, but also to classify firms by estimating the probability of each firm being constrained. The classification based on the switching regression is likely to be more reliable than the simple ex ante division based on one firm characteristic. Based on the firm classification, the percentage of the dispersion of MRPK that can be explained by the presence of constrained firms can be estimated.

In the switching regression model shown below, the probability of firm i being financially constrained or unconstrained is determined by the following selection equation:

$$s_{i,t}^{*} = \boldsymbol{x_{S,i,t}} \boldsymbol{\gamma^{S}} + \varepsilon_{S,i,t} \tag{4.26}$$

where $s_{i,t}^{*}$ is a latent variable and $\boldsymbol{x_{S,i,t}}$ is a row vector that contains variables affecting a firm's constrained status, including $\text{Age}_{i,t}$, size measured by $\text{lnAssets}_{i,t-1}$, $\text{lnMRPK}_{i,t-1}$, inverse leverage measured by the net-worth-to-assets ratio $\left(\frac{\text{Shareholders' Funds}}{\text{Assets}}\right)_{i,t-1}$, and

[34]Beck et al. (2005) find that the smallest (largest) firms are affected the most (least) by financing obstacles, using survey data from 54 countries in the late 1990s.

[35]For instance, using an indicator variable that equals one if age (or total assets) is below its 30th or 50th percentile, or if age is below an absolute value (e.g., age < 15 years), to indicate constrained firms yields inconclusive results. The results are only consistent with the hypothesis that constrained firm's investment is more sensitive to cash flow in around ten or fewer countries.

liquidity measured by cash-to-assets ratio $\left(\frac{\text{Cash}}{\text{Assets}}\right)_{i,t-1}$. In addition, vector $x_{S,i,t}$ also includes year dummies and four-digit industry dummies. The parameter $\boldsymbol{\gamma^S}$ is a column vector of the corresponding coefficients for the variables in $\boldsymbol{x_{S,i,t}}$. The error term $\varepsilon_{S,i,t}$ has a logistic distribution with mean zero and variance σ_S^2.

Constrained and unconstrained firms follow two different investment regimes, as the investment of the constrained firms should be more sensitive to fluctuations in internal financing. Which investment regime a firm follows depends on the selection equation (4.26). More specifically, the investment of a firm i follows the constrained (C) regime, i.e., $\Delta\text{lnFTA}_{i,t} = \Delta\text{lnFTA}_{i,t}^C$, if the latent variable $s_{i,t}^*$ is positive:

$$\Delta\text{lnFTA}_{i,t}^C = \boldsymbol{x_{C,i,t}\gamma^C} + \varepsilon_{C,i,t} \quad \text{if} \quad \boldsymbol{x_{S,i,t}\gamma^S} + \varepsilon_{S,i,t} > 0 \tag{4.27}$$

But it follows the unconstrained (U) regime, i.e., $\Delta\text{ln}k_{i,t} = \Delta\text{ln}k_{i,t}^U$, if the latent variable $s_{i,t}^*$ is nonpositive:

$$\Delta\text{lnFTA}_{i,t}^U = \boldsymbol{x_{U,i,t}\gamma^U} + \varepsilon_{U,i,t} \quad \text{if} \quad \boldsymbol{x_{S,i,t}\gamma^S} + \varepsilon_{S,i,t} \leqslant 0 \tag{4.28}$$

where $\boldsymbol{x_{C,i,t}}$ and $\boldsymbol{x_{U,i,t}}$ both contain the investment opportunity $\text{Opp}_{i,t-1}$, the availability of internal financing, year dummies, and four-digit industry dummies. As discussed in Section 4.4.1, $\text{Opp}_{i,t-1}$ is proxied by lagged sales growth, lagged value added growth, or lagged productivity growth, and the availability of internal financing is proxied by lagged cash flow over twice lagged capital stock. $\boldsymbol{\gamma^C}$ and $\boldsymbol{\gamma^U}$ are the corresponding coefficients for the variables in vectors $\boldsymbol{x_{C,i,t}}$ and $\boldsymbol{x_{U,i,t}}$ respectively. The differences between the parameters $\boldsymbol{\gamma^C}$ and $\boldsymbol{\gamma^U}$ reflect the differential investment behavior of firms in the two regimes. The error terms are normally distributed with mean zero and standard deviations of σ_C and σ_U, i.e., $\varepsilon_{C,i,t} \sim N(0, \sigma_C^2)$ and $\varepsilon_{U,i,t} \sim N(0, \sigma_U^2)$, where $\varepsilon_{C,i,t}$ and $\varepsilon_{U,i,t}$ are independent of $\varepsilon_{S,i,t}$. This paper uses an exogenous switching model as in Garcia et al. (1997), because firms do not choose to become constrained or unconstrained. However, if shocks to firms' investment are correlated with shocks to the financial variables in the selection equation (4.26), then an endogenous switching regression model, where $\varepsilon_{C,i,t}$ and $\varepsilon_{U,i,t}$ are each allowed to be correlated with the error term in the selection equation $\varepsilon_{S,i,t}$, may be more appropriate, as studied in Almeida and Campello (2007), Hovakimian and Titman (2006), and Hu and Schiantarelli (1998), for example.

Although theoretically, the availability of internal financing only matters for the constrained firms, as shown in (4.24), it is included in both investment regimes since (4.21) and (4.24) only hold if a firm's constrained status stays the same in two consecutive periods.[36] Furthermore, cash flow (as a proxy for the availability in internal financing) captures current profitability which may capture information about the investment opportunity that is not captured by the proxies such as lagged sales growth or productivity

[36]For example, if a firm is constrained in period $t-1$ but becomes unconstrained in period t, then its investment in period t still depends on its lagged net worth.

growth, so it can be significant in both regimes. However, assuming that the proxies such as lagged sales growth or productivity growth capture the investment opportunity equally well for the two types of firms, the presence of financial frictions would be expected to lead to a larger coefficient on cash flow and a lower coefficient on sales growth in the constrained regime.

The switching regression itself does not automatically identify which investment regime is associated with constrained firms. The identification of the constrained investment regime requires theoretical priors on how certain firm characteristics indicate firms' constrained status. Since both $\boldsymbol{x_{C,i,t}}$ and $\boldsymbol{x_{U,i,t}}$ contain the same variables, the constrained regime is identified using some of the variables included in $\boldsymbol{x_{S,i,t}}$. In this paper, the identification relies on the signs and the significance of the coefficients for age, size and MRPK, since it is relatively unambiguous that firms with younger age, smaller size and higher MRPK are more likely to be constrained. More specifically, if the coefficients for these three variables are each significant at 10% level at least, then the regime is classified as constrained if the probability for being in this regime increases in MRPK and decreases in age and size at the same time. If not all three variables are each significant at 10% level at least, then the regime classification relies only on the signs of the significant variables.

Although firm leverage and liquidity ratios are also included in the selection equation, they are not used for the identification of the constrained regime because a priori, it is ambiguous how firm leverage and liquidity indicate a firm's constrained status. One the one hand, firms with higher leverage could be more constrained as they have lower net worth and thus lower borrowing capacity. On the other hand, they may be unconstrained because the fact that they have higher leverage could mean they are able to borrow a lot in the first place. Similarly, firms with high liquidity measured by cash over total assets can be unconstrained if it indicates the firms are profitable and mature. However, it can indicate that firms are constrained if firms cannot easily borrow from the credit market and thus hold more cash as precautionary savings.

Although the financially constrained status of a firm is unobservable, according to (4.26) and (4.27), the probability of a firm i being constrained in period t can be specified as:

$$P(\boldsymbol{x_{S,i,t}\gamma^S} + \varepsilon_{S,i,t} > 0) = P(\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S}) \tag{4.29}$$

Assuming the error term $\varepsilon_{S,i,t}$ has a logistic distribution with mean zero and standard deviation of σ_S, i.e., $\varepsilon_{S,i,t} \sim Logit(0, \sigma_S^2)$, then the probability of firm i being constrained in period t is determined by a logit function:[37]

$$P(\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S}) = P(\varepsilon_{S,i,t} < \boldsymbol{x_{S,i,t}\gamma^S}) = \frac{\exp(\boldsymbol{x_{S,i,t}\gamma^S})}{1+\exp(\boldsymbol{x_{S,i,t}\gamma^S})} \tag{4.30}$$

The likelihood function of an observation $L_{i,t}$ is the weighted sum of the likelihoods of being in each latent class (i.e., the constrained and unconstrained groups of firms), where the weights are the associated latent class probabilities, $P(\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S})$ and $P(\varepsilon_{S,i,t} \leqslant -\boldsymbol{x_{S,i,t}\gamma^S})$. It is shown in Appendix C.5 that:

$$L_{i,t} = f(\varepsilon_{C,i,t})P(\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S}) + f(\varepsilon_{U,i,t})P(\varepsilon_{S,i,t} \leqslant -\boldsymbol{x_{S,i,t}\gamma^S}) \tag{4.31}$$

where $f(.)$ is the marginal normal density. It follows from (4.31) that the log-likelihood function for all observations is:

$$L = \sum_{i=1}^{M}\sum_{t=1}^{T_i}\ln(L_{i,t}) \tag{4.32}$$

where M is the number of firms in each industry and T_i is the number of observations for each firm i. By maximizing the log-likelihood function (4.32), the parameters γ^S, γ^C, γ^U, $\ln\sigma_C$, and $\ln\sigma_U$ can be estimated. With the estimated parameters, it is possible to calculate the posterior probability of each firm being in each of the two regimes. Once the regimes are identified as constrained or unconstrained, the posterior probability of a firm being in the constrained regime can be used to classify firms as constrained or unconstrained for each period t. In this paper, if the posterior probability of a firm being constrained is greater than 0.5 in any period, then this firm is classified as constrained in that period. The posterior probability of a firm being constrained $P(\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S}|\Delta\text{lnFTA}_{i,t})$ takes into account the information about investment by updating the prior probability based on Bayes' rule:

$$\frac{f(\varepsilon_{C,i,t}|\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S})P(\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S})}{f(\varepsilon_{C,i,t}|\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S})P(\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S}) + f(\varepsilon_{U,i,t}|\varepsilon_{S,i,t} \leqslant -\boldsymbol{x_{S,i,t}\gamma^S})P(\varepsilon_{S,i,t} \leqslant -\boldsymbol{x_{S,i,t}\gamma^S})} \tag{4.33}$$

where $P(\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S})$ is the prior probability for being constrained.

To control for unobserved firm heterogeneity, it would be desirable to add firm fixed effects, but there are two difficulties with doing so in a switching regression. Although adding firm dummies in all three equations (4.26)-(4.28) would control for the firm fixed effects, this is computationally very costly. For instance, if there are 1000 firms, then adding 1000 dummies into each of the three equations would result in 3000 additional

[37]Similarly, according to (4.26) and (4.28), the probability of firm i being unconstrained is:

$$P(\boldsymbol{x_{S,i,t}\gamma^S} + \varepsilon_{S,i,t} \leqslant 0) = P(\varepsilon_{S,i,t} \leqslant -\boldsymbol{x_{S,i,t}\gamma^S}) = 1 - P(\varepsilon_{S,i,t} \leqslant \boldsymbol{x_{S,i,t}\gamma^S}) \quad = \frac{1}{1+\exp(\boldsymbol{x_{S,i,t}\gamma^S})}$$

where the last step uses (4.30).

parameters to be estimated. Demeaning the two investment equations would impose a rather strong assumption on the nature of the firm fixed effects, i.e., firm heterogeneity in the two different regimes has the same impact on firm investment.

The second difficulty is that even if it is computationally feasible to include thousands of firm dummies, the 'incidental parameters problem' (Neyman and Scott, 1948) still exists in the nonlinear selection equation (4.26). This is because there are only T_i observations to estimate each firm i's dummy and the estimate of the firm dummy remains random even as the number of firms N grows. This randomness cannot be averaged out due to the nonlinearity, unlike in a linear model. Hence, in a nonlinear model with firm fixed effects and a fixed time dimension, the maximum likelihood estimators of the firm dummies and the explanatory variables are inconsistent (Greene, 2004; Chamberlain, 1980). Within transformation or first-differencing will not eliminate the individual firm heterogeneity in a nonlinear model either.[38]

Hu and Schiantarelli (1998) deal with the firm-fixed effects by modelling them as a linear function of the means of the firm-specific variables in each investment equation and the selection equation. They control for the means of these variables in each equation. Hovakimian and Titman (2006) adopt a different approach to partially control for firm fixed effects. They include the firm-specific variables and their lags in each equation and also the lagged dependent variable in each investment equation before estimating the switching regression. However, first differencing will not eliminate the unobserved firm heterogeneity in a nonlinear model. So this paper follows the approach used in Hu and Schiantarelli (1998) to control for firm fixed effects.

To further reduce the problem of unobserved firm heterogeneity, this paper applies the switching regression model (4.26)-(4.28) to each two-digit industry in each country. The existing literature that adopts the switching regression often uses US data only and runs the switching regression on a country level after controlling for two-digit industry fixed effects (e.g., Hu and Schiantarelli, 1998). Since this paper uses much larger datasets, where countries have a large number of firms even at a two-digit industry level, it is possible to run the switching regression at a more disaggregated level and then control for four-digit industry fixed effects in order to mitigate the problem of unobserved firm heterogeneity.[39] Restricting the sample to a two-digit industry also improves the reliability of the proxies used for MRPK (i.e., nominal revenue or nominal value added over fixed tangible assets, treating the revenue elasticity as a constant within a subsector), as it overcomes the

[38] In a few cases, including the logistic regression, the incidental parameters problem can be solved by conditioning on a sufficient statistic for the incidental parameters. For instance, the sufficient statistic is $\sum_{t=1}^{T_i} s_{i,t}$ for a logit model, where $s_{i,t}$ is the dependent variable that takes a value of zero or one. However, this conditional maximum likelihood approach cannot be used here, because whether a firm is constrained or not is unobserved (i.e., the value of $s_{i,t}$ is unknown).

[39] As can be seen in Table 4.1, there are more than 200 four-digit industries in nearly all of the countries in the baseline sample. If the switching regression model were run on the country level, then it would be infeasible to control for four-digit industry fixed effects for the reasons discussed above.

problem that the revenue elasticities β_k and β_m are likely to differ significantly across two-digit industries.

Using lagged sales growth to proxy for investment opportunity and lagged cash flow over twice lagged fixed tangible assets $\frac{\text{CF}_{i,t-1}}{\text{FTA}_{i,t-2}}$ to proxy for the availability of internal financing, Table 4.5 and 4.6 show the results from fitting the switching regression model (4.26)-(4.28) to the fabricated metal products industry (two-digit NACE Rev.2 code = 25) for each country separately. This industry has the largest number of observations.[40] Four-digit sector dummies and year dummies are included in each investment equation and the selection equation. To address the firm fixed effects, mean lagged sales growth and mean cash flow for each firm over time are included in the investment equations. The means of firm-specific variables (apart from age due to collinearity) for each firm over time in the selection equation are added as additional variables in the selection equation.

The switching regression itself does not automatically identify which investment regime is associated with constrained firms. The identification of the constrained regime relies on the theoretical priors that firms with higher MRPK, smaller size and younger age are more likely to be constrained. Although more variables are included in the selection equation, ex ante, variables such as leverage and liquidity ratios are ambiguous indicators for the constrained status, as discussed before in this section. Let P^C denote the probability of a firm being constrained, then rearranging (4.30) gives the log odds ratio $\ln\frac{P^C}{1-P^C} = \boldsymbol{x_{S,i,t}\gamma^S}$. As can be seen from this odds ratio, if the coefficient is positive, it means that when its corresponding variable increases, the probability of being constrained P^C also increases. Hence, if the probability of being in regime C increases in MRPK, but decreases in size and age, then regime C is the constrained regime.

Table 4.5 shows the coefficients for the key variables in the selection equation that determine whether a given regime is for the constrained firms (4.29). As can be seen, lagged log of MRPK is positive and significant in all countries, meaning that a higher MRPK increases the probability of a firm being in the constrained regime. Similarly, coefficients for age and size (proxied by log of total assets) are negative and highly significant in all countries except for one, meaning that an older age and a larger size will reduce (raise) the probability of a firm being constrained (unconstrained). This result is in line with the findings in Beck et al. (2005) that the smallest (largest) firms are affected the most (least) by financing obstacles. The coefficients for inverse leverage ratio measured by the shareholders' funds (net worth) to assets ratio have mixed signs across countries, implying that leverage ratio is an ambiguous indicator for a firm's constrained status,

[40] When summing or taking the mean of the number of observations across countries for a given industry, industry 25 has the highest data availability. On average, industry 25 accounts for around 17% of the total number of firm-year observations in the manufacturing sector in each country, while this percentage for the other industries is below 11%. The full category of the two-digit manufacturing industries and their descriptions can be found in Table C.5 in Appendix C.4. Industries such as the manufacture of tobacco products, coke and refined petroleum products, and basic pharmaceutical products are quite concentrated in the sample of 20 countries, so they do not have enough observations for the switching regression.

Table 4.5: The Selection Equation of the Switching Regression in Fabricated Metal Products Industry

Country	Age	ln(Assets)	ln(MRPK)	$\frac{\text{Net worth}}{\text{Assets}}$	$\frac{\text{Cash}}{\text{Assets}}$	Fraction constrained
Bulgaria	-0.024***	-0.846***	0.656***	0.581	1.949***	0.39
	(0.0063)	(0.2097)	(0.1116)	(0.3773)	(0.7238)	
Croatia	-0.021***	-0.871***	1.027***	-0.117	0.981**	0.41
	(0.0049)	(0.0901)	(0.0576)	(0.2039)	(0.4581)	
Czech Republic	-0.080***	-0.842***	0.921***	0.268*	1.912***	0.41
	(0.0047)	(0.0770)	(0.0479)	(0.1601)	(0.2661)	
Finland	-0.019***	-0.959***	1.375***	0.692***	1.155***	0.23
	(0.0022)	(0.0626)	(0.0582)	(0.1434)	(0.2074)	
France	-0.014***	-1.094***	2.089***	1.043***	1.884***	0.24
	(0.0009)	(0.0412)	(0.0297)	(0.0936)	(0.1150)	
Germany	-0.006***	-0.942***	1.585***	1.619***	1.466**	0.26
	(0.0013)	(0.2625)	(0.1432)	(0.4629)	(0.6148)	
Italy	-0.013***	-0.845***	1.084***	-0.354***	1.179***	0.34
	(0.0007)	(0.0254)	(0.0156)	(0.0768)	(0.1046)	
Japan	-0.016***	-0.579	1.388***	-0.179	1.607	0.25
	(0.0032)	(0.3534)	(0.2184)	(0.7774)	(1.0320)	
Korea	-0.026***	-0.729***	1.163***	0.571***	1.389***	0.36
	(0.0024)	(0.0465)	(0.0370)	(0.1379)	(0.2789)	
Norway	-0.017***	-0.608***	1.016***	0.093	0.626**	0.34
	(0.0050)	(0.0945)	(0.0572)	(0.2121)	(0.2795)	
Poland	-0.031***	-1.074***	1.129***	0.310	2.135***	0.27
	(0.0062)	(0.1130)	(0.0856)	(0.2258)	(0.4190)	
Portugal	-0.031***	-0.931***	1.315***	0.506***	1.329***	0.37
	(0.0019)	(0.0626)	(0.0400)	(0.1223)	(0.2196)	
Romania	-0.020***	-0.507***	0.820***	-0.114**	1.370***	0.44
	(0.0036)	(0.0322)	(0.0299)	(0.0542)	(0.1720)	
Serbia	-0.005	-0.517***	0.833***	-0.507***	2.362***	0.31
	(0.0040)	(0.0625)	(0.0512)	(0.1870)	(0.5014)	
Slovakia	-0.066***	-1.290***	1.092***	0.394**	1.780***	0.40
	(0.0076)	(0.1388)	(0.0851)	(0.1859)	(0.4017)	
Slovenia	-0.039***	-1.565***	1.350***	0.989***	3.177***	0.32
	(0.0055)	(0.1349)	(0.0949)	(0.3471)	(0.5918)	
Spain	-0.016***	-0.667***	1.289***	0.080	0.984***	0.30
	(0.0010)	(0.0241)	(0.0167)	(0.0544)	(0.0921)	
Sweden	-0.013***	-0.947***	1.331***	0.780***	1.103***	0.32
	(0.0013)	(0.0488)	(0.0353)	(0.1310)	(0.1456)	
Ukraine	-0.017***	-0.448***	0.605***	-0.103	0.575**	0.33
	(0.0033)	(0.0419)	(0.0325)	(0.0731)	(0.2805)	
United Kingdom	-0.005***	-0.650***	1.029***	0.229	1.008***	0.24
	(0.0012)	(0.0799)	(0.0617)	(0.1547)	(0.2658)	

Note: The table shows the coefficients for the key variables in the selection equation that determines the probability of a firm being constrained, including age, log of assets, log of MRPK, net worth-to assets ratio, and cash-to-assets ratio, and the average proportion of constrained firms over the sample period. All variables apart from age are lagged. Four-digit industry and year fixed effects are included. To address firm fixed effects, the means of firm-specific variables (apart from age) are controlled in the selection equation. The last column shows the average proportion of constrained firms over the sample period, where firms are classified as constrained based on the estimated posterior probabilities. Robust standard errors are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

as expected. Higher liquidity proxied by cash-to-assets ratio increases the probability of being constrained, reflecting that firms hold more cash for precautionary reasons due to the lack of easy access to external financing.

Table 4.5 also shows the average proportion of constrained firms over the sample period in the last column, where firms are classified as constrained based on the estimated posterior probability. Using the posterior probability as shown in (4.33), a firm is classified as constrained if the posterior probability of the firm being in the constrained regime is greater than 0.5 and otherwise, the firm is classified as unconstrained. The probabilities of being constrained and unconstrained sum to one for each firm. This average proportion of constrained firms across time for each country can also be seen in graph (a) in Figure 4.3.

In Table 4.6, the coefficients for the lagged sales growth and the cash flow from two different investment regimes, the constrained regime (4.27) and the unconstrained regime (4.28), are reported. As can be seen, the coefficient for cash flow is significant and much larger for constrained firms in all countries, whereas it is not significant for unconstrained firms in 12 out of 20 countries. As discussed before, since cash flow captures the current profitability and is positively related to expected future profitability due to persistence of profitability, it can be significant even for unconstrained firms. This explains why in 6 out of 20 countries, the coefficient on cash flow is highly significant for unconstrained firms. However, the coefficient on cash flow is much larger for constrained firms in all countries, so the results are consistent with the hypothesis that constrained firms' investment should be more sensitive to cash flow. In Romania, the coefficient on cash flow for unconstrained firms is negative and significant at 10% level, which is likely due to the unobserved firm heterogeneity not being fully controlled. The results without the firm fixed effects can be found in Table C.11 and C.12 in Appendix C.6, which show that the coefficient on cash flow for unconstrained firms can be negative without firm fixed effects.

The coefficient for lagged sales growth is significantly positive for unconstrained firms in all countries except for Bulgaria, whereas it is not significant for constrained firms in 7 out of 20 countries. However, in 12 countries, the coefficient on lagged sales growth for constrained firms is highly significant and larger than that for unconstrained firms, which differs from the theoretical prediction. This is likely because lagged sales growth is a poor proxy for investment opportunity, so the results on lagged sales growth are not very reliable. If there were a perfect measure for investment opportunity, then testing whether unconstrained firms' investment is more responsive to investment opportunity would be more meaningful.

I use a likelihood ratio test to test whether the switching regression model (less restrictive) fits the data better than a single regime model estimated by OLS (more restrictive). Following Hu and Schiantarelli (1998), the degrees of freedom equal the number of constraints (that the coefficients in the two investment equations are equal) plus the number of parameters in the selection equation, which are shown in the last

Table 4.6: Switching Regression Model of Firm Investment in Fabricated Metal Products Industry

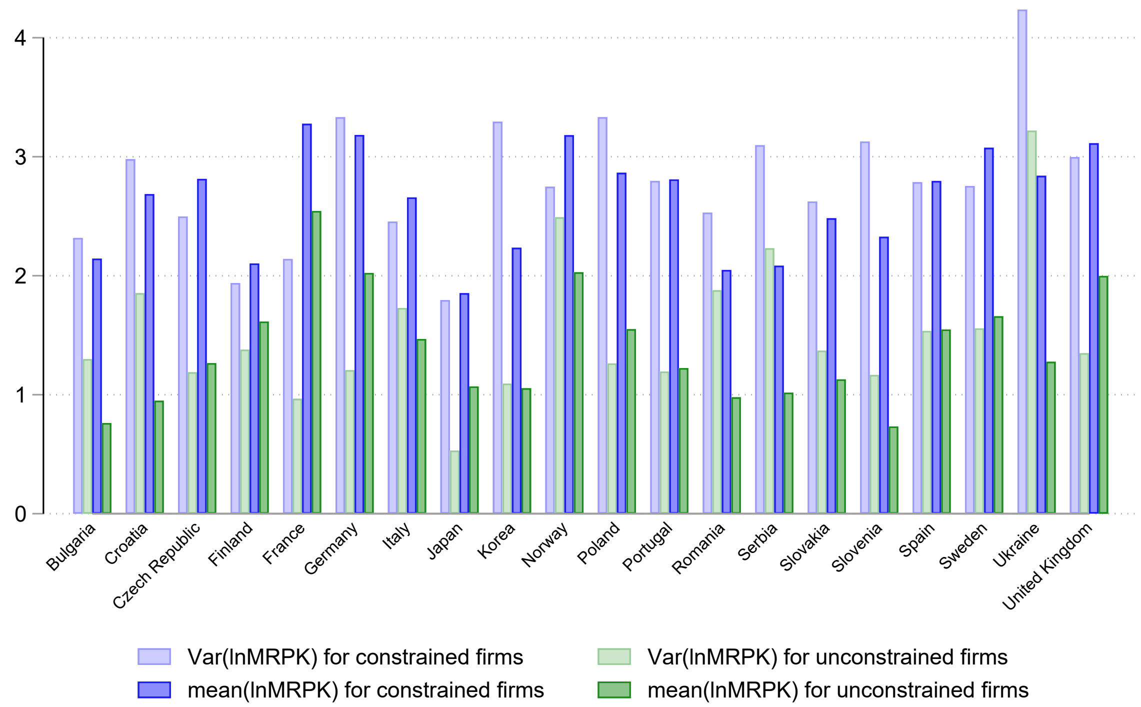
	Unconstrained Regime		Constrained Regime				
Country	$\Delta\text{lnSales}_{i,t-1}$	$\frac{\text{CF}_{i,t-1}}{\text{FTA}_{i,t-2}}$	$\Delta\text{lnSales}_{i,t-1}$	$\frac{\text{CF}_{i,t-1}}{\text{FTA}_{i,t-2}}$	Observations	Prob > Chi2	df
Bulgaria	0.010	0.007*	0.050	0.025***	4,243	0.0000	76
	(0.0060)	(0.0042)	(0.0410)	(0.0066)			
Croatia	0.014**	-0.002	0.156***	0.025***	12,652	0.0000	76
	(0.0057)	(0.0015)	(0.0277)	(0.0038)			
Czech Republic	0.021***	-0.002	0.023	0.031***	25,421	0.0000	96
	(0.0043)	(0.0015)	(0.0249)	(0.0028)			
Finland	0.024***	-0.000	0.007	0.033***	27,429	0.0000	80
	(0.0039)	(0.0011)	(0.0257)	(0.0047)			
France	0.111***	0.010***	0.124***	0.039***	170,850	0.0000	82
	(0.0048)	(0.0006)	(0.0217)	(0.0015)			
Germany	0.067***	0.003***	0.073	0.018***	12,100	0.0000	98
	(0.0099)	(0.0011)	(0.0501)	(0.0041)			
Italy	0.020***	0.002***	0.123***	0.045***	246,989	0.0000	94
	(0.0015)	(0.0007)	(0.0083)	(0.0017)			
Japan	0.013**	0.016	0.024	0.061*	6,830	0.0000	86
	(0.0063)	(0.0103)	(0.0546)	(0.0332)			
Korea	0.009***	0.000	0.064***	0.059***	55,900	0.0000	60
	(0.0026)	(0.0011)	(0.0194)	(0.0039)			
Norway	0.028***	0.001	0.136***	0.028***	12,676	0.0000	78
	(0.0078)	(0.0007)	(0.0434)	(0.0028)			
Poland	0.036***	0.004***	0.023	0.022***	13,237	0.0000	82
	(0.0061)	(0.0016)	(0.0416)	(0.0038)			
Portugal	0.025***	-0.001	0.151***	0.022***	47,373	0.0000	76
	(0.0037)	(0.0017)	(0.0191)	(0.0023)			
Romania	0.015***	-0.001*	0.046***	0.021***	44,863	0.0000	82
	(0.0023)	(0.0005)	(0.0106)	(0.0014)			
Serbia	0.021***	0.003***	0.099***	0.047***	12,866	0.0000	74
	(0.0035)	(0.0012)	(0.0235)	(0.0056)			
Slovakia	0.030***	-0.004	-0.041	0.040***	10,806	0.0000	82
	(0.0069)	(0.0032)	(0.0328)	(0.0064)			
Slovenia	0.020***	0.002	0.075*	0.047***	12,476	0.0000	66
	(0.0071)	(0.0025)	(0.0419)	(0.0063)			
Spain	0.023***	0.001	0.131***	0.042***	193,141	0.0000	84
	(0.0016)	(0.0005)	(0.0124)	(0.0017)			
Sweden	0.054***	0.003***	0.110***	0.031***	56,662	0.0000	78
	(0.0040)	(0.0007)	(0.0206)	(0.0020)			
Ukraine	0.011***	0.002	0.100***	0.009***	20,782	0.0000	70
	(0.0023)	(0.0011)	(0.0148)	(0.0023)			
United Kingdom	0.038***	0.001	0.071**	0.013***	26,117	0.0000	82
	(0.0059)	(0.0008)	(0.0326)	(0.0020)			

Note: The dependent variable is firm investment $\Delta\text{lnFTA}_{i,t}$. The coefficients for lagged sales growth and lagged cash flow in two different investment regimes are reported. Four-digit industry and year fixed effects are included in the switching regression. Firm fixed effects are partially controlled by adding the means of the firm-specific variables in each equation, whose coefficients are not reported here. The last two columns show the p-value for the likelihood ratio test and the degrees of freedom for the χ^2 distribution respectively. A small p-value suggests that the switching regression (less restrictive model) fits the data significantly better than an OLS regression. Robust standard errors are reported in parentheses.

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

column of Table 4.6. The right-tail p-values of the chi-squared statistic are also reported in Table 4.6.[41] The small p-values reject the null and suggest that the switching regression model fits the data significantly better than the single regime model.

Figure 4.2: Dispersions and Means of Marginal Revenue Product of Capital in Fabricated Metal Products Industry



Note: The bar chart shows the cross-section variances (or dispersions) and means of the ln(MRPK) for constrained firms and unconstrained firms in industry 25 (manufacture of fabricated metal products) by NACE Rev.2 Code across 20 countries. The dispersions and means of MRPK are averaged over the sample period. Constrained and unconstrained firms are identified using the results from the switching regression model. MRPK is computed as the nominal revenue divided by tangible fixed assets.
Data source: Orbis

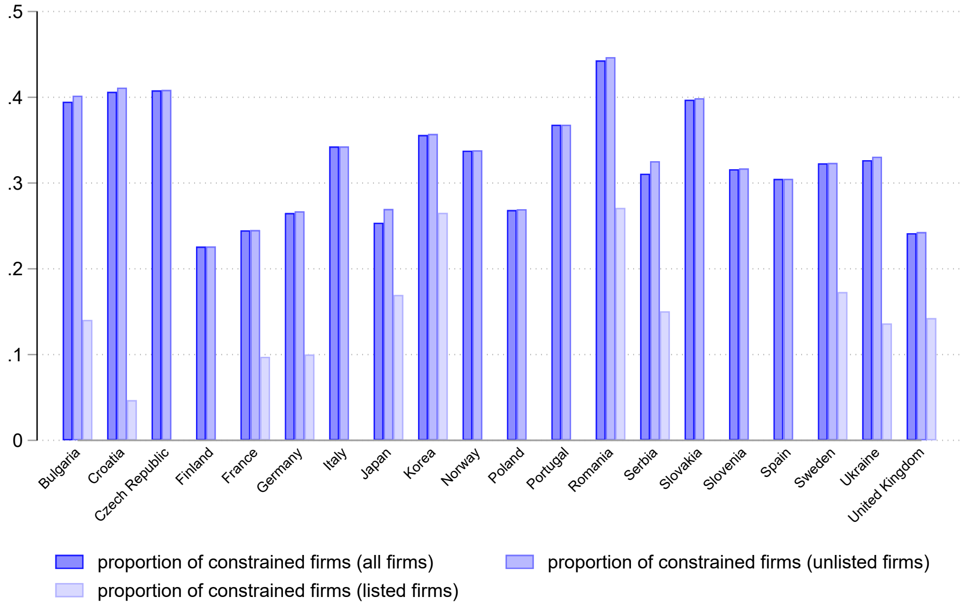
Figure 4.2 plots the cross-section variances (or dispersions) and means of ln(MRPK) for each type of firms in the fabricated metal products industry in each country, where the dispersions and means of ln(MRPK) are averaged over the sample period for each country. It can be seen that the dispersions and means of MRPK for constrained firms are much larger than those for unconstrained firms. These patterns are similar to those shown in Figure 4.1, although the contrast between the two types of firms is much more notable here. It is discussed in Section 4.2.3 that according to the model, constrained firms are expected to have a higher cross-section variance and mean of MRPK than unconstrained firms.

Figure 4.3 shows the proportion of constrained firms and credit distortion in the fabricated metal products industry (industry 25 by NACE Rev.2 Code), which are

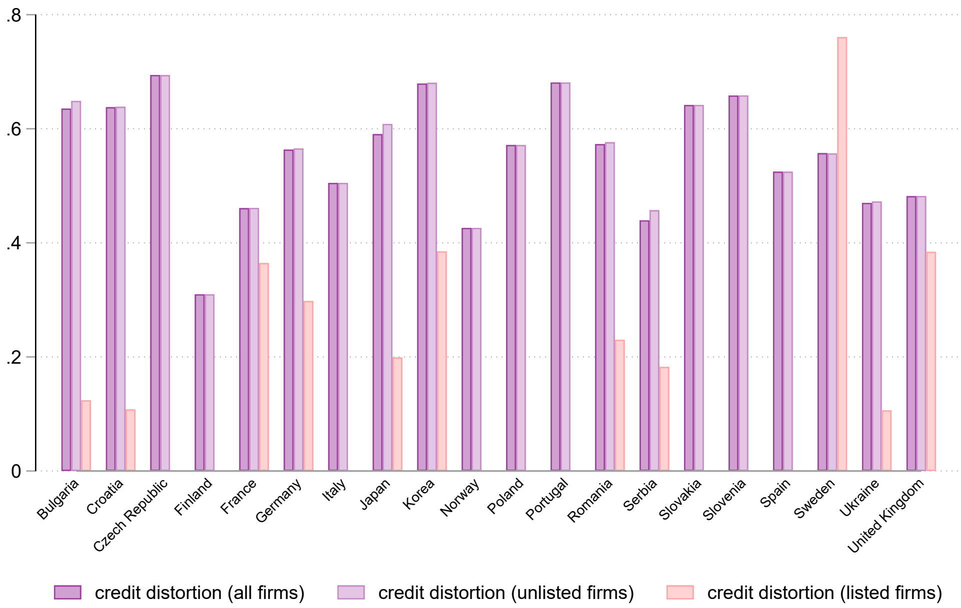
[41] The chi-squared statistic equals 2*(log likelihood of less restrictive model - log likelihood of more restrictive model).

Figure 4.3: Proportion of Constrained Firms and Credit Distortion in Fabricated Metal Products Industry

(a) Proportion of Constrained Firms



(b) Credit Distortion



Note: In each graph, the corresponding measure is computed across all firms in a given sample and all years. Three different samples of firms are used: all firms, unlisted firms and listed firms. The missing bars for the listed firms are due to the number of observations being below 100 over the sample period. Graph (a) plots the fraction of constrained firms in industry 25 (manufacture of fabricated metal products) by NACE Rev.2 Code across 20 countries. Graph (b) plots the credit distortion in percent points (i.e., the proportion of the observed dispersion (cross-section variance) of MRPK that is caused by the presence of constrained firms) in industry 25, which is computed based on (4.19). MRPK is computed as the nominal revenue divided by tangible fixed assets.
Data source: Orbis

averaged over the sample period across all firms in a given sample (i.e., entire sample, listed firms only, or unlisted firms only). Graph (a) shows that in most countries, the average proportion of constrained firms across all firms and years is above 0.25 and this proportion is much larger than the average proportion computed using the subsample of listed firms over the sample period. For example, in the UK, there are around 25% of firms classified as financially constrained across all firms and years, but less than 15% of listed firms are classified as constrained on average across all listed firms and years. This is consistent with the expectation that large listed firms are less likely to be financially constrained.

The missing bars for the listed firms in some countries in Figure 4.3 are because the number of observations for listed firms is below 100 over all years, in which case the sample of listed firms is not representative and the two measures are not computed. Fewer than 100 observations for the listed firms over the sample period implies that there are around ten listed firms per year on average, given that most countries have more than ten years of data, as can be seen from Table 4.1. Furthermore, graph (a) in Figure 4.3 shows that the proportion of constrained firms is slightly above 0.1 in many countries, implying that only one out of the ten listed firms is classified as constrained, so this sample is not representative.

Classifying firms into the constrained and unconstrained groups is only the first step. The main objective of the paper is to estimate the impact of the financial constraints on the dispersion of MRPK. Based on (4.19), the fraction of the dispersion of MRPK that is caused by the presence of constrained firms can be computed. Graph (b) in Figure 4.3 plots the average credit distortion (in percent points) in the fabricated metal products industry, which are averaged over the sample period across all firms in a given sample. As shown in graph (b), the credit distortion ranges from 0.3 in Finland to around 0.7 in Czech Republic, Korea and Portugal, which means the presence of constrained firms in this industry can explain around 30-70% of the dispersion of MRPK, depending on different countries. Note that the average credit distortion computed using the subsample of listed firms tends to be lower except for Sweden.

I apply the switching regression model in the baseline analysis to all the other industries that have enough observations to run the switching regression.[42] The results for 14 different industries (including industry 25 in the baseline analysis) are summarized in Table C.7 and C.8 in Appendix C.6. Each column in Table C.7 summarizes the average proportion of constrained firms across all firms and years in a given industry for different countries. The column for industry 25 in Table C.7 corresponds to the height of the first bar in graph (a) in Figure 4.3. In addition, the last two rows of Table C.7 show the mean

[42]There are nine two-digit industries that do not have enough observations for most countries (i.e., industry 11, 12, 15, 17, 19, 21, 24, 29, and 30 by NACE Rev.2 code). Industry 33 (repair and installation of machinery and equipment) is neglected since it is not a typical manufacturing industry. Industry 33 only accounts for around 4.5% of the total number of firm-year observations in the manufacturing sector in each country on average.

difference between the proportion of constrained firms in a sample of unlisted firms and that in a sample of listed firms, and the number of countries used to calculate this mean difference, respectively. For example, this mean difference across 11 countries is 0.18 for industry 25 (fabricated metal products industry), as shown in Table C.7.

Similarly, each column in Table C.8 summarizes the credit distortion across all firms and years in a given industry for different countries. The column for industry 25 in Table C.8 corresponds to the height of the first bar in graph (b) in Figure 4.3. The last two rows of Table C.8 show the mean difference between the credit distortion in a sample of unlisted firms and the credit distortion in a sample of listed firms, and the number of countries used to compute this mean difference, respectively.

As can be seen from Table C.7 and C.8, a general pattern is that for most two-digit industries and countries, there is at least a quarter of firms being classified as financially constrained and the presence of these firms explain more than half of the dispersion of MRPK across all firms. In addition, for each reported industry, the proportion of constrained firms within unlisted firms is larger than that within listed firms by around 0.15 except for industry 32 and the credit distortion within unlisted firms is higher by around 0.2 except for industries 26 and 32 on average across countries.

For robustness checks, I also use nominal value added to compute MRPK and the results for the proportion of constrained firms and credit distortion are summarized in Table C.9 and C.10 in Appendix C.6, which are very similar to the baseline results. In addition, I compare the baseline results with the case without trying to use the Hu and Schiantarelli (1998) approach to control for firm fixed effects, so with only four-digit industry and year dummies included. The coefficients for the investment equation and the selection equation can be found in Table C.11 and C.12 in Appendix C.6. The negative coefficients on cash flow for unconstrained firms are likely caused by the between variation in cash flow when firm fixed effects are ignored. Despite the differences in the coefficients, the results for the proportion of constrained firms and credit distortion are robust to whether firm fixed effects are included, as shown in Figure C.1. Finally, I use different proxies for investment opportunity, i.e., lagged value added growth and productivity growth. The results for the proportion of constrained firms and credit distortion are robust to whether lagged sales growth, value added growth or productivity growth is used, except for Korea, as shown in Figure C.2.

4.5 Conclusions

This paper proposes a novel method for estimating the impact of financial frictions on capital misallocation measured by the dispersion of the marginal revenue product of capital (MRPK), which uses large firm-level datasets and requires fewer restrictive assumptions. The key idea is that the observed dispersion of MRPK can be viewed in terms of the dispersions and means for the financially constrained and unconstrained firms. Based on the decomposition of the dispersion of MRPK, this paper provides a credit distortion measure, which measures the proportion of the observed dispersion of MRPK that can be attributed to the presence of constrained firms.

A simple model of firm dynamics with a one-period time to build for capital and borrowing constraints shows that the capital decisions and thus the MRPK for constrained and unconstrained firms are driven by two different processes. While the capital investment of an unconstrained firm is driven by the expected future investment opportunity, that of a constrained firm is determined by the availability of its internal financing. As a result, the distribution of MRPK across all firms is a mixture of two distributions, one for the constrained firms and the other for the unconstrained firms.

By decomposing the dispersion of MRPK across all firms into the dispersions and means of MRPK for the two types of firms, this paper provides new insights into the mechanisms through which the borrowing constraint increases the dispersion of MRPK. While the usual mechanism is through a higher MRPK of constrained firms relative to the unconstrained firms, an often neglected mechanism is through the dispersions within the constrained and unconstrained firms.

More importantly, this statistical decomposition provides a way to estimate the impact of financial frictions on the dispersion of MRPK once the constrained firms are identified. Using a switching regression model to identify the constrained firms in the manufacturing industry for 20 countries from the 1990s to 2015, this paper finds that the dispersions and means of MRPK within the group of constrained firms are much larger than those within the group of unconstrained firms, which are consistent with the model predictions. Furthermore, for most two-digit industries and countries, more than a quarter of firms are classified as financially constrained and the presence of these constrained firms accounts for more than half of the dispersion of MRPK.

Therefore, this paper has quantified the impact of financially constrained firms on the allocation of physical capital.

Appendix A

Appendix to Chapter 2

A.1 Solving the Entrepreneur's Problem

The proof resembles the approach used by Andrés and Arce (2012) in solving for c_t^E and b_t. Substitute $\lambda_{1,t}^E = \frac{1}{c_t^E}$ (2.15) and $\lambda_{2,t}^E = \frac{1}{c_t^E} - \beta^E E_t \left[\frac{1}{c_{t+1}^E} \frac{R_{b,t}}{\pi_{t+1}} \right]$ (2.16) into the first order condition (2.18) with respect to k_t^E and rearrange:

$$\frac{q_t - m_{k,t} E_t \left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}} \right]}{c_t^E} = \beta^E E_t \left[\frac{1}{c_{t+1}^E} \left\{ \frac{\alpha y_{w,t+1}}{x_{t+1} k_t^E} + q_{t+1}(1-\delta) - \frac{m_{k,t}}{\pi_{t+1}} E_t[q_{t+1}(1-\delta)\pi_{t+1}] \right\} \right] \tag{A.1}$$

Multiply both sides by k_t^E:

$$\begin{aligned} &\frac{q_t k_t^E - m_{k,t} E_t \left[\frac{q_{t+1} k_t^E (1-\delta)\pi_{t+1}}{R_{b,t}} \right]}{c_t^E} \\ =&\beta^E E_t \left[\frac{1}{c_{t+1}^E} \left\{ \frac{\alpha y_{w,t+1}}{x_{t+1}} + q_{t+1} k_t^E (1-\delta) - \frac{m_{k,t}}{\pi_{t+1}} E_t[q_{t+1} k_t^E (1-\delta)\pi_{t+1}] \right\} \right] \end{aligned} \tag{A.2}$$

Similarly, substitute the expressions for $\lambda_{1,t}^E$ and $\lambda_{2,t}^E$ into the first order condition (2.19) with respect to h_t^E and rearrange:

$$\frac{q_{h,t} - m_{h,t} E_t \left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}} \right]}{c_t^E} = \beta^E E_t \left[\frac{1}{c_{t+1}^E} \left\{ \frac{\upsilon y_{w,t+1}}{x_{t+1} h_t^E} + q_{h,t+1} - \frac{m_{h,t}}{\pi_{t+1}} E_t[q_{h,t+1}\pi_{t+1}] \right\} \right] \tag{A.3}$$

Multiply both sides by h_t^E:

$$\frac{q_{h,t} h_t^E - m_{h,t} E_t \left[\frac{q_{h,t+1} h_t^E \pi_{t+1}}{R_{b,t}} \right]}{c_t^E} = \beta^E E_t \left[\frac{1}{c_{t+1}^E} \left\{ \frac{\upsilon y_{w,t+1}}{x_{t+1}} + q_{h,t+1} h_t^E - \frac{m_{h,t}}{\pi_{t+1}} E_t[q_{h,t+1} h_t^E \pi_{t+1}] \right\} \right] \tag{A.4}$$

Add equations (A.2) and (A.4), and substitute the binding borrowing constraint (2.23) to simplify:

$$\frac{q_t k_t^E + q_{h,t} h_t^E - b_t}{c_t^E} = \beta^E E_t \left[\frac{1}{c_{t+1}^E} \left\{ \frac{(\alpha+v) y_{w,t+1}}{x_{t+1}} + q_{t+1} k_t^E (1-\delta) + q_{h,t+1} h_t^E - \frac{R_{b,t} b_t}{\pi_{t+1}} \right\} \right] \tag{A.5}$$

Recall the definition for the entrepreneur's net worth n_t and the rewritten budget constraint:

$$n_t \equiv \frac{(\alpha+v) y_{w,t}}{x_t} + q_t (1-\delta) k_{t-1}^E + q_{h,t} h_{t-1}^E - \frac{R_{b,t-1} b_{t-1}}{\pi_t} \tag{2.21}$$

$$c_t^E + q_t k_t^E + q_{h,t} h_t^E = n_t + b_t \tag{2.22}$$

Using (2.21) and (2.22), equation (A.5) can be written as:

$$\frac{n_t - c_t^E}{c_t^E} = \beta^E E_t \left[\frac{n_{t+1}}{c_{t+1}^E} \right] \tag{A.6}$$

Conjecture that $c_t^E = \gamma n_t$ and substitute the conjecture into the equation above:

$$\frac{(1-\gamma) n_t}{\gamma n_t} = \beta^E E_t \left[\frac{n_{t+1}}{\gamma n_{t+1}} \right] \tag{A.7}$$

Hence, $\gamma = (1 - \beta^E)$, so $c_t^E = (1-\beta^E) n_t$ (2.24) and $b_t = q_t k_t^E + q_{h,t} h_t^E - \beta^E n_t$ (2.25).

A.2 Calvo Pricing

A.2.1 Optimal Pricing Equation

Substitute in $y_{t+s}^*(j)$ and rearrange:

$$Max_{p_t^*(j)} \quad \sum_{s=0}^{\infty} \theta^s E_t \left[\Lambda_{t,t+s} \left(\frac{p_t^*(j)}{p_{t+s}} - \frac{1}{x_{t+s}} \right) \left(\frac{p_t^*(j)}{p_{t+s}} \right)^{-\epsilon} y_{t+s} \right] \tag{A.8}$$

Take the first order condition:

$$\sum_{s=0}^{\infty} \theta^s E_t \Lambda_{t,t+s} \left[\left(\frac{1}{p_{t+s}} \right) \left(\frac{p_t^*(j)}{p_{t+s}} \right)^{-\epsilon} y_{t+s} + \left(\frac{p_t^*(j)}{p_{t+s}} - \frac{1}{x_{t+s}} \right) (-\epsilon) \left(\frac{p_t^*(j)}{p_{t+s}} \right)^{-\epsilon-1} \frac{y_{t+s}}{p_{t+s}} \right] = 0 \tag{A.9}$$

Simplify the above equation:

$$\sum_{s=0}^{\infty} \theta^s E_t \Lambda_{t,t+s} \left[(1-\epsilon) \left(\frac{y_{t+s}}{p_{t+s}} \right) \left(\frac{p_t^*(j)}{p_{t+s}} \right)^{-\epsilon} + \epsilon \frac{1}{x_{t+s}} p_t^*(j)^{-\epsilon-1} \left(\frac{1}{p_{t+s}} \right)^{-\epsilon} y_{t+s} \right] = 0 \tag{A.10}$$

Multiply by $\frac{p_t^*(j)^{\epsilon+1}}{1-\epsilon}$:

$$\sum_{s=0}^{\infty} \theta^s E_t \Lambda_{t,t+s} \left[p_t^*(j) \left(\frac{1}{p_{t+s}} \right)^{1-\epsilon} y_{t+s} + \frac{\epsilon}{1-\epsilon} \frac{1}{x_{t+s}} \left(\frac{1}{p_{t+s}} \right)^{-\epsilon} y_{t+s} \right] = 0 \tag{A.11}$$

Rearrange to solve for $p_t^*(j)$ and get the optimal pricing equation (2.35):

$$p_t^*(j) = \frac{\epsilon}{\epsilon - 1} \frac{\sum_{s=0}^{\infty} \theta^s E_t \left[\Lambda_{t,t+s} x_{t+s}^{-1} p_{t+s}^{\epsilon} y_{t+s} \right]}{\sum_{s=0}^{\infty} \theta^s E_t \left[\Lambda_{t,t+s} p_{t+s}^{\epsilon-1} y_{t+s} \right]} = \frac{\epsilon}{\epsilon - 1} \frac{\sum_{s=0}^{\infty} (\beta\theta)^s E_t \left[u'(c_{t+s}) x_{t+s}^{-1} p_{t+s}^{\epsilon} y_{t+s} \right]}{\sum_{s=0}^{\infty} (\beta\theta)^s E_t \left[u'(c_{t+s}) p_{t+s}^{\epsilon-1} y_{t+s} \right]} \tag{A.12}$$

To numerically implement the optimal pricing equation in Dynare, summarize the equation above with two recursive formulations such that:

$$p_t^* = \frac{\epsilon}{\epsilon - 1} \frac{g_{1,t}}{g_{2,t}} \tag{A.13}$$

where

$$g_{1,t} \equiv u'(c_t) p_t^{\epsilon} y_t x_t^{-1} + \beta\theta E_t[g_{1,t+1}] = \frac{1}{c_t} p_t^{\epsilon} y_t x_t^{-1} + \beta\theta E_t[g_{1,t+1}] \tag{A.14}$$

$$g_{2,t} \equiv u'(c_t) p_t^{\epsilon-1} y_t + \beta\theta E_t[g_{2,t+1}] = \frac{1}{c_t} p_t^{\epsilon-1} y_t + \beta\theta E_t[g_{2,t+1}] \tag{A.15}$$

Let $f_{1,t} \equiv p_t^{-\epsilon} g_{1,t}$, then

$$f_{1,t} \equiv p_t^{-\epsilon} g_{1,t} = \frac{1}{c_t} y_t x_t^{-1} + \beta\theta E_t[\pi_{t+1}^{\epsilon} f_{1,t+1}] \tag{A.16}$$

Let $f_{2,t} \equiv p_t^{1-\epsilon} g_{2,t}$, then

$$f_{2,t} \equiv p_t^{1-\epsilon} g_{2,t} = \frac{1}{c_t} y_t + \beta\theta E_t[\pi_{t+1}^{\epsilon-1} f_{2,t+1}] \tag{A.17}$$

The optimal pricing equation $p_t^* = \frac{\epsilon}{\epsilon-1} \frac{g_{1,t}}{g_{2,t}}$ becomes:

$$p_t^* = \frac{\epsilon}{\epsilon - 1} \frac{f_{1,t} p_t^{\epsilon}}{f_{2,t} p_t^{\epsilon-1}} = \frac{\epsilon}{\epsilon - 1} \frac{f_{1,t}}{f_{2,t}} p_t \tag{A.18}$$

Divide both sides by p_{t-1} and let $\pi_t^* = \frac{p_t^*}{p_{t-1}}$ denote the gross reset price inflation rate to eliminate the price levels:

$$\pi_t^* = \frac{p_t^*}{p_{t-1}} = \frac{\epsilon}{\epsilon - 1} \frac{f_{1,t}}{f_{2,t}} \pi_t \tag{A.19}$$

A.2.2 Aggregate Price Evolution

Rearrange the aggregate price index (30):

$$p_t^{1-\epsilon} = \int_0^1 p_t(j)^{1-\epsilon} dj \tag{A.20}$$

Following Sims (2014),[1] the above integral can be broken up into two parts by ordering the retailers along the unit interval:

$$p_t^{1-\epsilon} = \int_0^{1-\theta} (p_t^*)^{1-\epsilon} dj + \int_{1-\theta}^{1} p_{t-1}(j)^{1-\epsilon} dj = (1-\theta)(p_t^*)^{1-\epsilon} + \int_{1-\theta}^{1} p_{t-1}(j)^{1-\epsilon} dj \quad \text{(A.21)}$$

Given the assumptions that the price-adjusting retailers in each period are randomly chosen and the number of retailers is large, the integral of individual prices over $[1-\theta, 1]$ of the unit interval is equal to a proportion θ of the integral over the entire unit interval, where θ is the length of the subset $[1-\theta, 1]$. That is,

$$\int_{1-\theta}^{1} p_{t-1}(j)^{1-\epsilon} dj = \theta \int_0^1 p_{t-1}(j)^{1-\epsilon} dj = \theta p_{t-1}^{1-\epsilon} \quad \text{(A.22)}$$

Hence, the aggregate price level evolves according to (2.36):

$$p_t^{1-\epsilon} = (1-\theta)(p_t^*)^{1-\epsilon} + \theta p_{t-1}^{1-\epsilon} \quad \text{(2.36)}$$

To compute the model numerically, it is necessary to rewrite the price evolution in terms of the inflation rates because the price level may not be stationary. Eliminating the price levels in the equation above by dividing both sides by $p_{t-1}^{1-\epsilon}$:

$$\left(\frac{p_t}{p_{t-1}}\right)^{1-\epsilon} = \theta + (1-\theta)\left(\frac{p_t^*}{p_{t-1}}\right)^{1-\epsilon} \quad \text{(A.23)}$$

Let $\pi_t \equiv \frac{p_t}{p_{t-1}}$ and $\pi_t^* \equiv \frac{p_t^*}{p_{t-1}}$ denote the gross inflation rate and the gross reset price inflation rate respectively, then the equation above can be rewritten as:

$$\pi_t^{1-\epsilon} = \theta + (1-\theta)(\pi_t^*)^{1-\epsilon} \quad \text{(A.24)}$$

A.2.3 Price Dispersion

Use the Calvo assumption to break up the integral into two parts by ordering the retailers along the unit interval:

$$f_{3,t} \equiv \int_0^1 \left[\frac{p_t(j)}{p_t}\right]^{-\epsilon} dj = \int_0^{1-\theta} \left(\frac{p_t^*}{p_t}\right)^{-\epsilon} dj + \int_{1-\theta}^{1} \left[\frac{p_{t-1}(j)}{p_t}\right]^{-\epsilon} dj \quad \text{(A.25)}$$

Rearrange and simplify by using the definitions for π_t and π_t^*:

$$f_{3,t} = \int_0^{1-\theta} \left(\frac{p_t^*}{p_{t-1}} \frac{p_{t-1}}{p_t}\right)^{-\epsilon} dj + \int_{1-\theta}^{1} \left[\frac{p_{t-1}(j)}{p_{t-1}} \frac{p_{t-1}}{p_t}\right]^{-\epsilon} dj = (1-\theta)(\pi_t^*)^{-\epsilon}\pi_t^{\epsilon} + \pi_t^{\epsilon} \int_{1-\theta}^{1} \left[\frac{p_{t-1}(j)}{p_{t-1}}\right]^{-\epsilon} dj \quad \text{(A.26)}$$

[1]https://www3.nd.edu/~esims1/new_keynesian_2014.pdf

Use the same method as in Appendix A.2.2 to simplify the last term in the equation above:

$$\int_{1-\theta}^{1}\left[\frac{p_{t-1}(j)}{p_{t-1}}\right]^{-\epsilon}dj=\theta\int_{0}^{1}\left[\frac{p_{t-1}(j)}{p_{t-1}}\right]^{-\epsilon}dj=\theta f_{3,t-1} \tag{A.27}$$

Hence, the price dispersion $f_{3,t}$ can be written recursively:

$$f_{3,t}=(1-\theta)(\pi_t^*)^{-\epsilon}\pi_t^{\epsilon}+\pi_t^{\epsilon}\theta f_{3,t-1} \tag{A.28}$$

As can be seen, the index j has been eliminated in the above expression. Consequently, there is no need to keep track of the individual prices. Using (2.37), (A.25) and (A.28), the final consumption good output y_t is:

$$y_t=\frac{y_{w,t}}{f_{3,t}}=\frac{y_{w,t}}{(1-\theta)(\pi_t^*)^{-\epsilon}\pi_t^{\epsilon}+\pi_t^{\epsilon}\theta f_{3,t-1}} \tag{A.29}$$

The real profit Π_t^R made by the continuum of unit mass retailers is:

$$\Pi_t^R=\int_0^1\left[\frac{p_t(j)}{p_t}y_t(j)-\frac{1}{x_t}y_t(j)\right]dj=\int_0^1\frac{p_t(j)}{p_t}y_t(j)dj-\frac{1}{x_t}\int_0^1 y_t(j)dj \tag{A.30}$$

Use retailer $j's$ individual demand function $y_t(j)=\left[\frac{p_t(j)}{p_t}\right]^{-\epsilon}y_t$ (2.31), the wholesale good output expression $y_{w,t}=\int_0^1 y_t(j)dj$ (2.37), the aggregate price index $p_t=\left[\int_0^1 p_t(j)^{1-\epsilon}dj\right]^{\frac{1}{1-\epsilon}}$ (2.32), and (A.29) to get (2.38):

$$\Pi_t^R=\int_0^1\frac{p_t(j)}{p_t}\left[\frac{p_t(j)}{p_t}\right]^{-\epsilon}y_t dj-\frac{y_{w,t}}{x_t}=y_t p_t^{\epsilon-1}\int_0^1 p_t(j)^{1-\epsilon}dj-\frac{y_{w,t}}{x_t}=y_t-\frac{y_{w,t}}{x_t}=\left(\frac{1}{f_{3,t}}-\frac{1}{x_t}\right)y_{w,t} \tag{2.38}$$

A.3 Elasticity of Loan Demand

A.3.1 Elasticities of Capital and Housing Demand to the Loan Rate

Use (2.8) to rewrite the first order condition (A.1) with respect to k_t^E as:

$$\begin{aligned}q_t-m_{k,t}E_t\left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}}\right]&=\beta^E E_t\left[\frac{c_t^E}{c_{t+1}^E}\frac{\alpha z_{t+1}(l_{t+1}^E)^{1-\alpha-v}}{x_{t+1}}\right](k_t^E)^{\alpha-1}(h_t^E)^v\\&+\beta^E E_t\left[\frac{c_t^E}{c_{t+1}^E}\left\{q_{t+1}(1-\delta)-\frac{m_{k,t}}{\pi_{t+1}}E_t[q_{t+1}(1-\delta)\pi_{t+1}]\right\}\right]\end{aligned} \tag{A.31}$$

Use notations $A_{k,t}$, $B_{k,t}$ and $C_{k,t}$ to simplify the above expression:

$$A_{k,t}\equiv q_t-m_{k,t}E_t\left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}}\right] \tag{A.32}$$

$$B_{k,t} \equiv \beta^E E_t \left[\frac{c_t^E}{c_{t+1}^E} \frac{\alpha z_{t+1} (l_{t+1}^E)^{1-\alpha-v}}{x_{t+1}} \right] > 0 \tag{A.33}$$

$$C_{k,t} \equiv \beta^E E_t \left[\frac{c_t^E}{c_{t+1}^E} \left\{ q_{t+1}(1-\delta) - \frac{m_{k,t}}{\pi_{t+1}} E_t[q_{t+1}(1-\delta)\pi_{t+1}] \right\} \right] \tag{A.34}$$

So (A.31) can be rewritten as:

$$A_{k,t} = B_{k,t} (k_t^E)^{\alpha-1} (h_t^E)^v + C_{k,t} \tag{A.35}$$

Similarly, rewrite the first order condition (A.3) with respect to h_t^E using the following notations:

$$A_{h,t} \equiv q_{h,t} - m_{h,t} E_t \left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}} \right] \tag{A.36}$$

$$B_{h,t} \equiv \beta^E E_t \left[\frac{c_t^E}{c_{t+1}^E} \frac{v z_{t+1} (l_{t+1}^E)^{1-\alpha-v}}{x_{t+1}} \right] > 0 \tag{A.37}$$

$$C_{h,t} \equiv \beta^E E_t \left[\frac{c_t^E}{c_{t+1}^E} \left\{ q_{h,t+1} - \frac{m_{h,t}}{\pi_{t+1}} E_t[q_{h,t+1}\pi_{t+1}] \right\} \right] \tag{A.38}$$

Hence, (A.3) can be written as:

$$A_{h,t} = B_{h,t} (k_t^E)^{\alpha} (h_t^E)^{v-1} + C_{h,t} \tag{A.39}$$

Rearrange (A.35) and substitute into the above equation:

$$A_{h,t} = B_{h,t} \left(\frac{A_{k,t} - C_{k,t}}{B_{k,t}} \right)^{\frac{\alpha}{\alpha-1}} (h_t^E)^{\frac{\alpha v}{1-\alpha}} (h_t^E)^{v-1} + C_{h,t} \tag{A.40}$$

Rearrange the equation above to solve for h_t^E:

$$h_t^E = \left(\frac{A_{h,t} - C_{h,t}}{B_{h,t}} \right)^{\frac{1-\alpha}{v-1+\alpha}} \left(\frac{A_{k,t} - C_{k,t}}{B_{k,t}} \right)^{\frac{\alpha}{v-1+\alpha}} = \left(\frac{A_{h,t} - C_{h,t}}{B_{h,t}} \right)^{u_1} \left(\frac{A_{k,t} - C_{k,t}}{B_{k,t}} \right)^{u_2} \tag{A.41}$$

where

$$u_1 \equiv -\frac{1-\alpha}{1-v-\alpha} < 0 \tag{A.42}$$

$$u_2 \equiv -\frac{\alpha}{1-v-\alpha} < 0 \tag{A.43}$$

Substitute the expression (A.41) for h_t^E into (A.35):

$$\begin{aligned} k_t^E &= \left(\frac{A_{k,t} - C_{k,t}}{B_{k,t}} \right)^{\frac{1}{\alpha-1}} \left(\frac{A_{h,t} - C_{h,t}}{B_{h,t}} \right)^{\frac{v}{v-1+\alpha}} \left(\frac{A_{k,t} - C_{k,t}}{B_{k,t}} \right)^{\frac{\alpha v}{(v-1+\alpha)(1-\alpha)}} \\ &= \left(\frac{A_{k,t} - C_{k,t}}{B_{k,t}} \right)^{u_3} \left(\frac{A_{h,t} - C_{h,t}}{B_{h,t}} \right)^{u_4} \end{aligned} \tag{A.44}$$

where

$$u_3 \equiv -\frac{1}{1-\alpha}\left(1+\frac{\alpha v}{1-v-\alpha}\right) = -\frac{1-v}{1-v-\alpha} < 0 \tag{A.45}$$

$$u_4 \equiv -\frac{v}{1-v-\alpha} < 0 \tag{A.46}$$

Note that $R_{b,t}$ is present in both $A_{h,t}$ and $A_{k,t}$. Use (A.41) and (A.44) to differentiate the two choice variables h_t^E and k_t^E with respect to $R_{b,t}$:

$$\begin{aligned}\frac{\partial h_t^E}{\partial R_{b,t}} =& u_1\left(\frac{A_{h,t}-C_{h,t}}{B_{h,t}}\right)^{u_1-1}\left(\frac{D_{h,t}}{B_{h,t}}\right)\left(\frac{A_{k,t}-C_{k,t}}{B_{k,t}}\right)^{u_2} \\ &+\left(\frac{A_{h,t}-C_{h,t}}{B_{h,t}}\right)^{u_1} u_2\left(\frac{A_{k,t}-C_{k,t}}{B_{k,t}}\right)^{u_2-1}\left(\frac{D_{k,t}}{B_{k,t}}\right) \\ =& \left(u_1\frac{D_{h,t}}{A_{h,t}-C_{h,t}}+u_2\frac{D_{k,t}}{A_{k,t}-C_{k,t}}\right)h_t^E < 0\end{aligned} \tag{A.47}$$

$$\begin{aligned}\frac{\partial k_t^E}{\partial R_{b,t}} =& u_4\left(\frac{A_{h,t}-C_{h,t}}{B_{h,t}}\right)^{u_4-1}\left(\frac{D_{h,t}}{B_{h,t}}\right)\left(\frac{A_{k,t}-C_{k,t}}{B_{k,t}}\right)^{u_3} \\ &+\left(\frac{A_{h,t}-C_{h,t}}{B_{h,t}}\right)^{u_4} u_3\left(\frac{A_{k,t}-C_{k,t}}{B_{k,t}}\right)^{u_3-1}\left(\frac{D_{k,t}}{B_{k,t}}\right) \\ =& \left(u_4\frac{D_{h,t}}{A_{h,t}-C_{h,t}}+u_3\frac{D_{k,t}}{A_{k,t}-C_{k,t}}\right)k_t^E < 0\end{aligned} \tag{A.48}$$

where

$$D_{h,t} \equiv m_{h,t}E_t\left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}^2}\right] > 0 \tag{A.49}$$

$$D_{k,t} \equiv m_{k,t}E_t\left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}^2}\right] > 0 \tag{A.50}$$

Note that it can be seen from (A.47) and (A.48) that $\frac{\partial h_t^E}{\partial R_{b,t}} < 0$ and $\frac{\partial k_t^E}{\partial R_{b,t}} < 0$ if $A_{h,t} - C_{h,t} > 0$ and $A_{k,t} - C_{k,t} > 0$. These sufficient conditions hold, as can be seen from (A.35) and (A.39). As a result, it follows from (2.50) that the market loan demand is downward-sloping: $\frac{\partial b_t}{\partial R_{b,t}} < 0$.

Note that this method of solving for $\frac{\partial h_t^E}{\partial R_{b,t}}$ and $\frac{\partial k_t^E}{\partial R_{b,t}}$ is the same as doing implicit differentiation in the two first order conditions, (A.1) and (A.3).

A.3.2 Elasticity of Loan Demand to the Loan Rate

To derive the elasticity PED_t of market loan demand to the loan rate, use (2.50) to get:

$$\begin{aligned} PED_t &\equiv -\frac{R_{b,t}}{b_t}\frac{\partial b_t}{\partial R_{b,t}} \\ &= -\frac{R_{b,t}}{b_t}\left\{-\frac{b_t}{R_{b,t}} + m_{h,t}E_t\left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}}\right]\frac{\partial h_t^E}{\partial R_{b,t}} + m_{k,t}E_t\left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}}\right]\frac{\partial k_t^E}{\partial R_{b,t}}\right\} \\ &= 1 - \frac{R_{b,t}}{b_t}m_{h,t}E_t\left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}}\right]\frac{\partial h_t^E}{\partial R_{b,t}} - \frac{R_{b,t}}{b_t}m_{k,t}E_t\left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}}\right]\frac{\partial k_t^E}{\partial R_{b,t}} \\ &= 1 - \frac{R_{b,t}}{b_t}\frac{b_{h,t}}{h_t^E}\frac{\partial h_t^E}{\partial R_{b,t}} - \frac{R_{b,t}}{b_t}\frac{b_{k,t}}{k_t^E}\frac{\partial k_t^E}{\partial R_{b,t}} \\ &= 1 + \frac{b_{h,t}}{b_t}PEH_t + \frac{b_{k,t}}{b_t}PEK_t > 0 \end{aligned} \tag{A.51}$$

where $PEH_t \equiv -\frac{\partial h_t^E}{\partial R_{b,t}}\frac{R_{b,t}}{h_t^E}$ and $PEK_t \equiv -\frac{\partial k_t^E}{\partial R_{b,t}}\frac{R_{b,t}}{k_t^E}$ denote the elasticities of housing and capital demand to the loan rate, respectively, and

$$b_{h,t} \equiv m_{h,t}E_t\left[\frac{q_{h,t+1}h_t^E\pi_{t+1}}{R_{b,t}}\right] > 0 \tag{A.52}$$

$$b_{k,t} \equiv m_{k,t}E_t\left[\frac{q_{t+1}k_t^E(1-\delta)\pi_{t+1}}{R_{b,t}}\right] > 0 \tag{A.53}$$

$$b_t \equiv b_{h,t} + b_{k,t} > 0 \tag{A.54}$$

The elasticity of loan demand to the loan rate is positive $PED_t > 0$ because the entrepreneur's demand for housing and physical capital decreases with the loan rate (i.e., $\frac{\partial h_t^E}{\partial R_{b,t}} < 0$ and $\frac{\partial k_t^E}{\partial R_{b,t}} < 0$), as shown in Appendix A.3.1.

To find $\frac{\partial h_t^E}{\partial R_{b,t}}\frac{R_{b,t}}{h_t^E}$, use the expressions for $\frac{\partial h_t^E}{\partial R_{b,t}}$ (A.47), (A.39), $B_{h,t}$ (A.37), $D_{h,t}$ (A.49), (A.35), $B_{k,t}$ (A.33), $D_{k,t}$ (A.50), $b_{h,t} \equiv m_{h,t}E_t\left[\frac{q_{h,t+1}h_t^E\pi_{t+1}}{R_{b,t}}\right]$ (A.52) and $b_{k,t} \equiv m_{k,t}E_t\left[\frac{q_{t+1}k_t^E(1-\delta)\pi_{t+1}}{R_{b,t}}\right]$ (A.53) to get:

$$\begin{aligned} \frac{\partial h_t^E}{\partial R_{b,t}}\frac{R_{b,t}}{h_t^E} &= \left(u_1\frac{D_{h,t}}{A_{h,t}-C_{h,t}} + u_2\frac{D_{k,t}}{A_{k,t}-C_{k,t}}\right)R_{b,t} \\ &= \left(u_1\frac{D_{h,t}}{B_{h,t}(k_t^E)^{\alpha}(h_t^E)^{v-1}} + u_2\frac{D_{k,t}}{B_{k,t}(k_t^E)^{\alpha-1}(h_t^E)^{v}}\right)R_{b,t} \\ &= \frac{u_1 m_{h,t}E_t\left[q_{h,t+1}\pi_{t+1}R_{b,t}^{-1}\right]}{\beta^E E_t\left[\frac{c_t^E}{c_{t+1}^E}\frac{v z_{t+1}(l_{t+1}^E)^{1-\alpha-v}}{x_{t+1}}\right](k_t^E)^{\alpha}(h_t^E)^{v-1}} + \frac{u_2 m_{k,t}E_t\left[q_{t+1}(1-\delta)\pi_{t+1}R_{b,t}^{-1}\right]}{\beta^E E_t\left[\frac{c_t^E}{c_{t+1}^E}\frac{\alpha z_{t+1}(l_{t+1}^E)^{1-\alpha-v}}{x_{t+1}}\right](k_t^E)^{\alpha-1}(h_t^E)^{v}} \\ &= \frac{u_1 m_{h,t}E_t\left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}}\right]}{E_t\left[\Lambda_{t,t+1}^E MPH_{t+1}\right]} + \frac{u_2 m_{k,t}E_t\left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}}\right]}{E_t\left[\Lambda_{t,t+1}^E MPK_{t+1}\right]} < 0 \end{aligned} \tag{A.55}$$

where $\Lambda^E_{t,t+1} \equiv \beta^E \frac{u'(c^E_{t+1})}{u'(c^E_t)} = \beta^E \frac{c^E_t}{c^E_{t+1}}$, $MPH_{t+1} \equiv \frac{v z_{t+1}(l^E_{t+1})^{1-\alpha-v}(k^E_t)^\alpha (h^E_t)^{v-1}}{x_{t+1}}$ and $MPK_{t+1} \equiv \frac{\alpha z_{t+1}(l^E_{t+1})^{1-\alpha-v}(k^E_t)^{\alpha-1}(h^E_t)^v}{x_{t+1}}$ denote the marginal products of housing and capital in terms of the final good, respectively. $\frac{\partial h^E_t}{\partial R_{b,t}} \frac{R_{b,t}}{h^E_t}$ is negative because $u_1 \equiv -\frac{1-\alpha}{1-v-\alpha} < 0$ and $u_2 \equiv -\frac{\alpha}{1-v-\alpha} < 0$.

Using (A.55) and the definitions for u_1 and u_2, the elasticity of the entrepreneur's housing demand to the loan rate PEH_t is:

$$PEH_t \equiv -\frac{\partial h^E_t}{\partial R_{b,t}} \frac{R_{b,t}}{h^E_t} = \frac{m_{h,t} E_t \left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}}\right] \frac{1-\alpha}{1-v-\alpha}}{E_t \left[\Lambda^E_{t,t+1} MPH_{t+1}\right]} + \frac{m_{k,t} E_t \left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}}\right] \frac{\alpha}{1-v-\alpha}}{E_t \left[\Lambda^E_{t,t+1} MPK_{t+1}\right]} > 0 \tag{A.56}$$

Similarly, use the expressions for $\frac{\partial k^E_t}{\partial R_{b,t}}$ (A.48), (A.39), $B_{h,t}$ (A.37), $D_{h,t}$ (A.49), (A.35), $B_{k,t}$ (A.33), $D_{k,t}$ (A.50), $b_{h,t} \equiv m_{h,t} E_t \left[\frac{q_{h,t+1} h^E_t \pi_{t+1}}{R_{b,t}}\right]$ (A.52) and $b_{k,t} \equiv m_{k,t} E_t \left[\frac{q_{t+1} k^E_t (1-\delta)\pi_{t+1}}{R_{b,t}}\right]$ (A.53) to get:

$$\frac{\partial k^E_t}{\partial R_{b,t}} \frac{R_{b,t}}{k^E_t} = \frac{u_4 m_{h,t} E_t \left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}}\right]}{E_t \left[\Lambda^E_{t,t+1} MPH_{t+1}\right]} + \frac{u_3 m_{k,t} E_t \left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}}\right]}{E_t \left[\Lambda^E_{t,t+1} MPK_{t+1}\right]} < 0 \tag{A.57}$$

where $u_4 \equiv -\frac{v}{1-v-\alpha} < 0$ and $u_3 \equiv -\frac{1-v}{1-v-\alpha} < 0$.

Using (A.57) and the definitions for u_4 and u_3, the elasticity of the entrepreneur's capital demand to the loan rate PEK_t is:

$$PEK_t \equiv -\frac{\partial k^E_t}{\partial R_{b,t}} \frac{R_{b,t}}{k^E_t} = \frac{m_{h,t} E_t \left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}}\right] \frac{v}{1-v-\alpha}}{E_t \left[\Lambda^E_{t,t+1} MPH_{t+1}\right]} + \frac{m_{k,t} E_t \left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}}\right] \frac{1-v}{1-v-\alpha}}{E_t \left[\Lambda^E_{t,t+1} MPK_{t+1}\right]} > 0 \tag{A.58}$$

As can be seen from (A.56) and (A.58), both PEH_t and PEK_t depend positively on the pledgeability ratios ($m_{h,t}$ and $m_{k,t}$) and the expected discounted values of the future prices of housing $E_t \left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}}\right]$ and capital $E_t \left[\frac{q_{t+1}\pi_{t+1}}{R_{b,t}}\right]$. Besides, they depend negatively on the expected discounted values of the marginal products of housing $E_t \left[\Lambda^E_{t,t+1} MPH_{t+1}\right]$ and capital $E_t \left[\Lambda^E_{t,t+1} MPK_{t+1}\right]$ in terms of the final good.

Substituting PEH_t (A.56) and PEK_t (A.58) into PED_t (A.51) gives:

$$\begin{aligned} PED_t =& 1 + \frac{m_{h,t} E_t \left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}}\right] \left(\frac{b_{h,t}}{b_t} \frac{1-\alpha}{1-v-\alpha} + \frac{b_{k,t}}{b_t} \frac{v}{1-v-\alpha}\right)}{E_t \left[\Lambda^E_{t,t+1} MPH_{t+1}\right]} \\ &+ \frac{m_{k,t} E_t \left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}}\right] \left(\frac{b_{h,t}}{b_t} \frac{\alpha}{1-v-\alpha} + \frac{b_{k,t}}{b_t} \frac{1-v}{1-v-\alpha}\right)}{E_t \left[\Lambda^E_{t,t+1} MPK_{t+1}\right]} \\ =& 1 + \frac{m_{h,t} E_t \left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}}\right] \left(\frac{b_{h,t}}{b_t} + \frac{v}{1-v-\alpha}\right)}{E_t \left[\Lambda^E_{t,t+1} MPH_{t+1}\right]} + \frac{m_{k,t} E_t \left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}}\right] \left(\frac{b_{k,t}}{b_t} + \frac{\alpha}{1-v-\alpha}\right)}{E_t \left[\Lambda^E_{t,t+1} MPK_{t+1}\right]} > 0 \end{aligned} \tag{A.59}$$

where the last step uses $b_t \equiv b_{h,t} + b_{k,t}$ (A.54). It can be seen from (A.59) that higher expected discounted values of marginal products of capital and housing and lower expected discounted values of prices of housing and capital (i.e., lower $E_t\left[\frac{q_{h,t+1}\pi_{t+1}}{R_{b,t}}\right]$ and $E_t\left[\frac{q_{t+1}(1-\delta)\pi_{t+1}}{R_{b,t}}\right]$) reduce the elasticity of the loan demand PED_t. Besides, a decrease in $m_{h,t}$ or $m_{k,t}$ (e.g., after a negative collateral shock) directly reduces PED_t, as shown in (A.59), and indirectly through raising the expected marginal products of capital and housing and lowering the expected asset prices.

A.4 Calibration and Steady State Values

Table A.1: Calibration of Parameters in Baseline Analysis

Parameter	Value	Description
Households		
β	0.995	Subjective discount factor
ϕ_l	1.45	Relative utility weight on leisure time
ϕ_h	0.1	Relative utility weight on housing
Entrepreneurs		
β^E	0.97	Subjective discount factor
α	0.33	Physical capital share
v	0.05	Housing share
δ	0.025	Depreciation rate for physical capital
Capital producers		
χ	10	Investment adjustment cost
Retailers		
ϵ	6	Elasticity of substitution between retail goods
θ	0.75	Probability of not adjusting price
Banking sector		
N	4	Number of banks
m_h	0.8	Loan-to-value ratio for housing
m_k	0.5	Loan-to-value ratio for physical capital
Central bank		
ρ_r	0.8	Interest rate smoothing
κ_π	1.5	Feedback coefficient on inflation
κ_y	0	Feedback coefficient on output

Table A.2: Steady State Values of Variables for Baseline Calibration

	Perfect Competition	Imperfect Competition
Gross Inflation Rate π	1	1
Productivity z	1	1
Housing Pledgeability Ratio m_h	0.5	0.8
Capital Pledgeability Ratio m_k	0.5	0.8
Output y	0.801	0.737
Consumption c	0.573	0.546
Investment i	0.127	0.109
Physical Capital k	5.081	4.358
Real Price of Capital q	1	1
Real Price of Housing q_h	14.694	12.896
Bank Loan b	5.042	3.668
Labor l	0.333	0.325
Real Wage w	1.244	1.172
Gross Real Deposit Rate R_r	1.005	1.005
Gross Real Loan Rate R_{rb}	1.005	1.011
Entrepreneur's Housing h^E	0.220	0.154
Entrepreneur's Net Worth n	3.378	2.754
Leverage Ratio $\frac{q_h h^E + qk}{n}$	2.463	2.302
Lagrange Multiplier λ_2^E	0.248	0.229
Loan Demand Elasticity PED	59.886	39.528
Capital Demand Elasticity PEK	24.089	20.607
Housing Demand Elasticity PEH	92.157	62.538

Note: The table shows the steady state values of selected variables under two types of banking competition: perfect banking competition and Cournot banking competition ($N = 4$). The steady state values for gross inflation rate, productivity and the two pledgeability ratios are exogenously set. The steady state values of all other variables are determined in equilibrium, based on the parameter values in Table A.1.

A.5 Dynare Model Block

Households

Household intertemporal consumption Euler equation (2.6)

Household intratemporal consumption-labor choice (2.4)

Household demand for real estate (2.5)

Marginal utility of the household's consumption (2.3)

Entrepreneurs

Entrepreneur's utility maximization with respect to k_t^E (2.18) and l_t^E (2.17)

Entrepreneur's borrowing $b_t = q_t k_t^E + q_{h,t} h_t^E - \beta^E n_t$ (2.25)

Binding borrowing constraint (2.23)

Wholesale good output $y_{w,t}$ (2.8)

Entrepreneur's consumption (2.24)

Entrepreneur's net worth (2.21)

Lagrange multiplier associated with the borrowing constraint (2.16)

Marginal utility of the entrepreneur's consumption (2.15)

AR(1) processes for log productivity (2.9), log pledgeability ratio for housing (2.13), and log pledgeability ratio for capital (2.14)

Retailers

$f_{1,t}$ (A.16) and $f_{2,t}$ (A.17) in optimal price rule

Gross reset price inflation rate (A.19)

Aggregate price evolution (written in terms of gross inflation rates) (2.36)

Recursive form of price dispersion $f_{3,t}$ (A.28)

Capital Producers

Real price of capital (2.29) and gross investment in physical capital (2.26)

Banking Sector

Perfectly competitive banks' profit maximization with respect to b_t (2.43)

Elasticity of loan demand to the loan rate $PED_t \equiv -\frac{\partial b_t}{\partial R_{b,t}} \frac{R_{b,t}}{b_t}$ (A.59)

Bank j's profit maximization with respect to $b_t(j)$ under Cournot competition (2.51)

Real loan rate $R_{rb,t} = E_t \left[\frac{R_{b,t}}{\pi_{t+1}} \right]$ and real deposit rate $R_{r,t}$ (2.45)

Real loan margin $RLM_t = R_{rb,t} - R_{r,t}$

Central Bank

Taylor rule (2.44)

Market Clearing

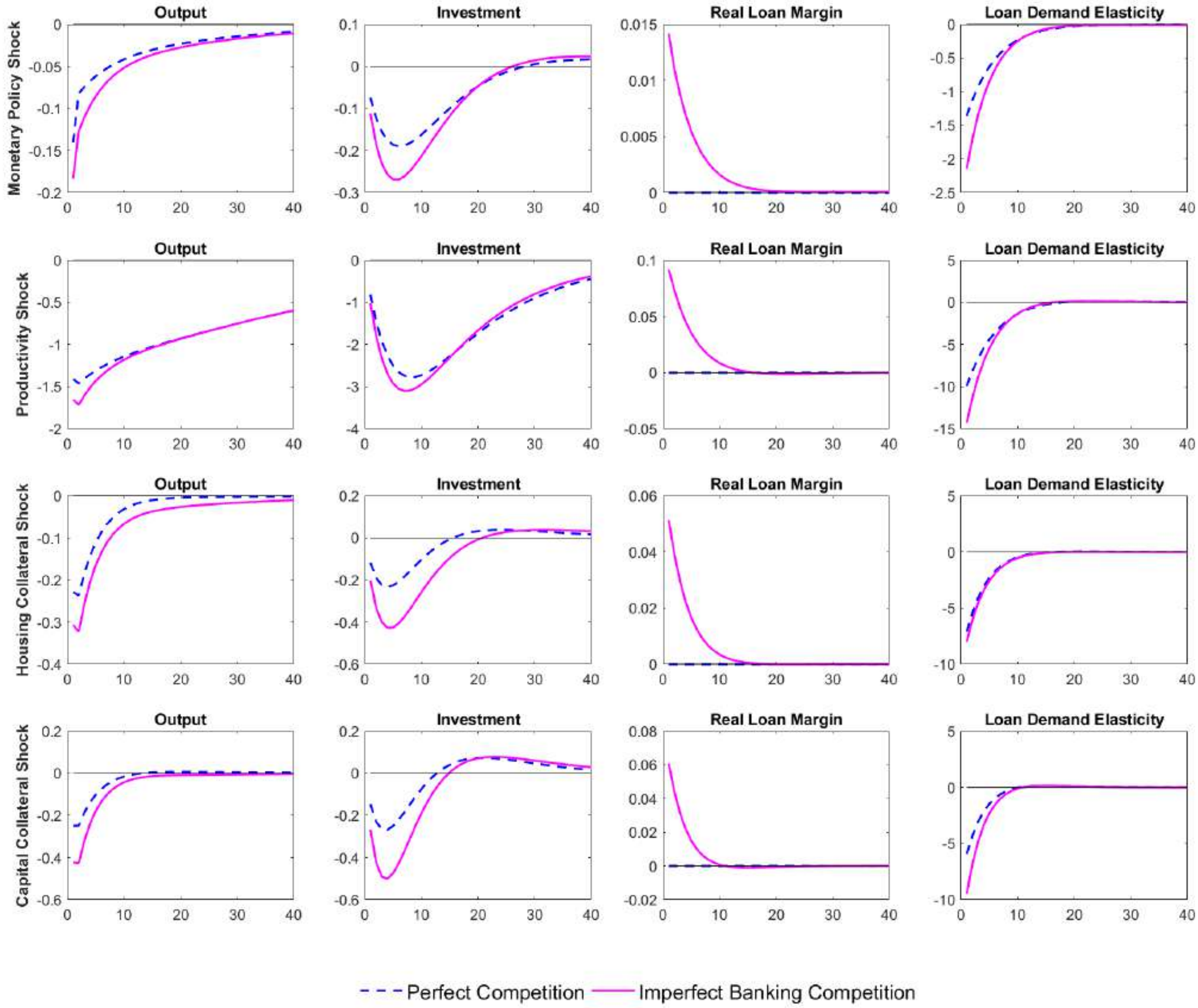
Final consumption good output y_t (A.29)

Aggregate resource constraint (2.52)

Housing market clearing condition $h_t + h_t^E = 1$ (total housing supply normalised to 1)

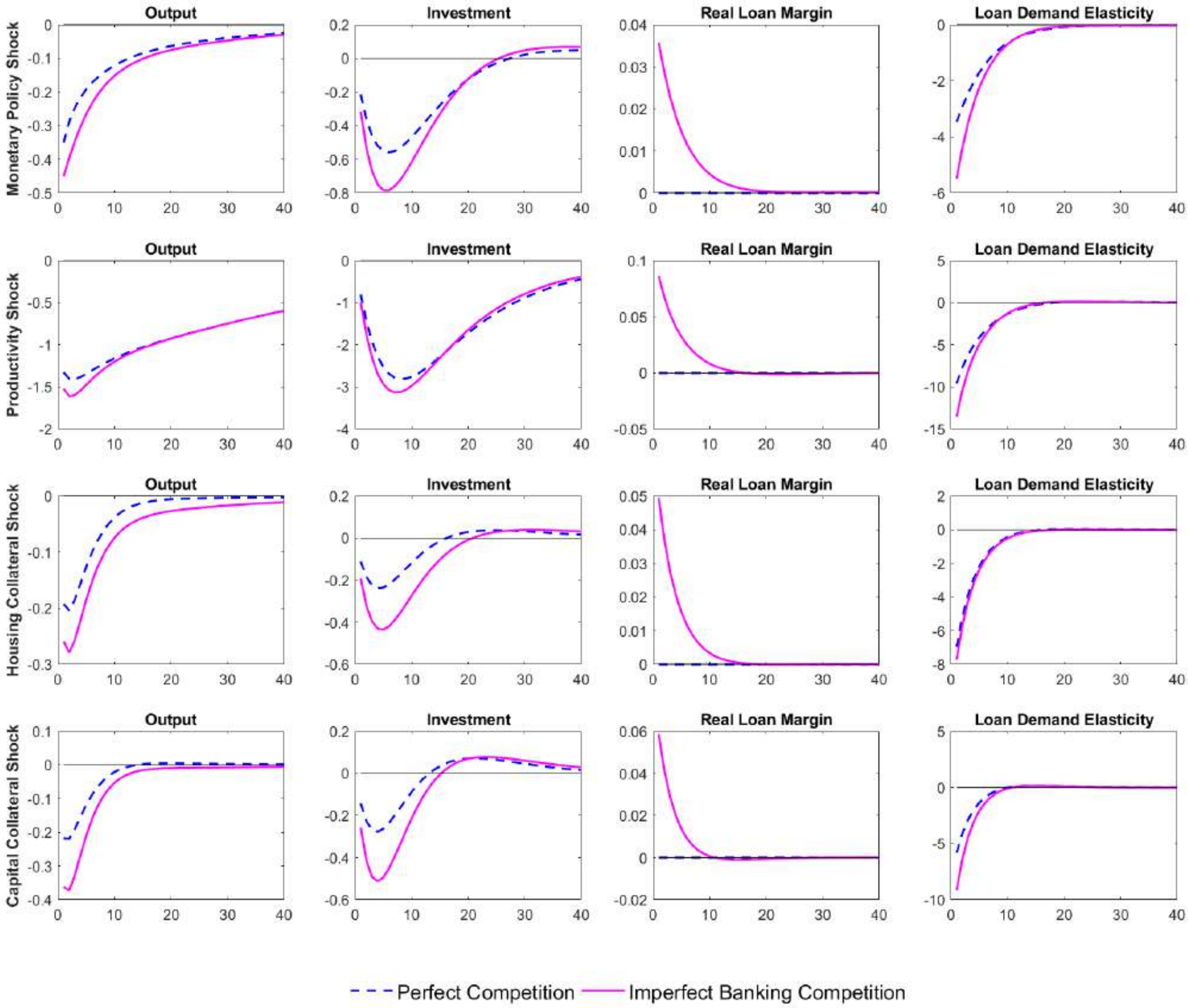
A.6 Robustness Checks

Figure A.1: Impulse Responses for Different Shocks without Interest Rate Smoothing ($\rho_r = 0$)



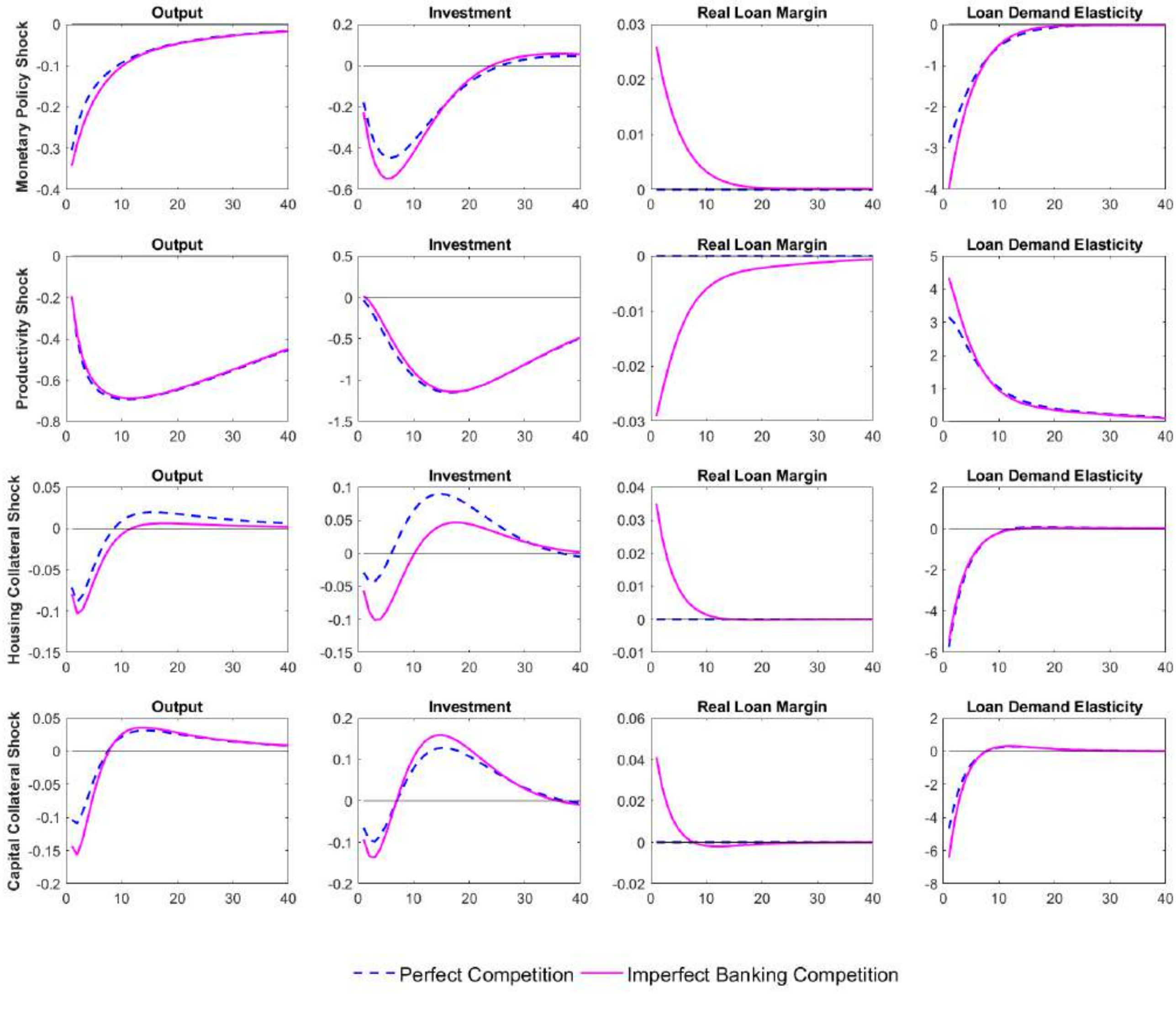
Note: Horizontal axis shows quarters after the shock that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state for variables other than the real loan margin, which is expressed in deviations from the steady state in percent points. The blue dashed line corresponds to perfect banking competition. The pink solid line corresponds to imperfect banking competition. Each row shows the impulse responses of four variables after a given type of shock: a 10 basis-point contractionary monetary policy shock, a one-standard-deviation negative productivity shock, and one-standard-deviation negative shocks to the pledgeability ratios m_h and m_k.

Figure A.2: Impulse Responses for Different Shocks when $\kappa_\pi = 3$



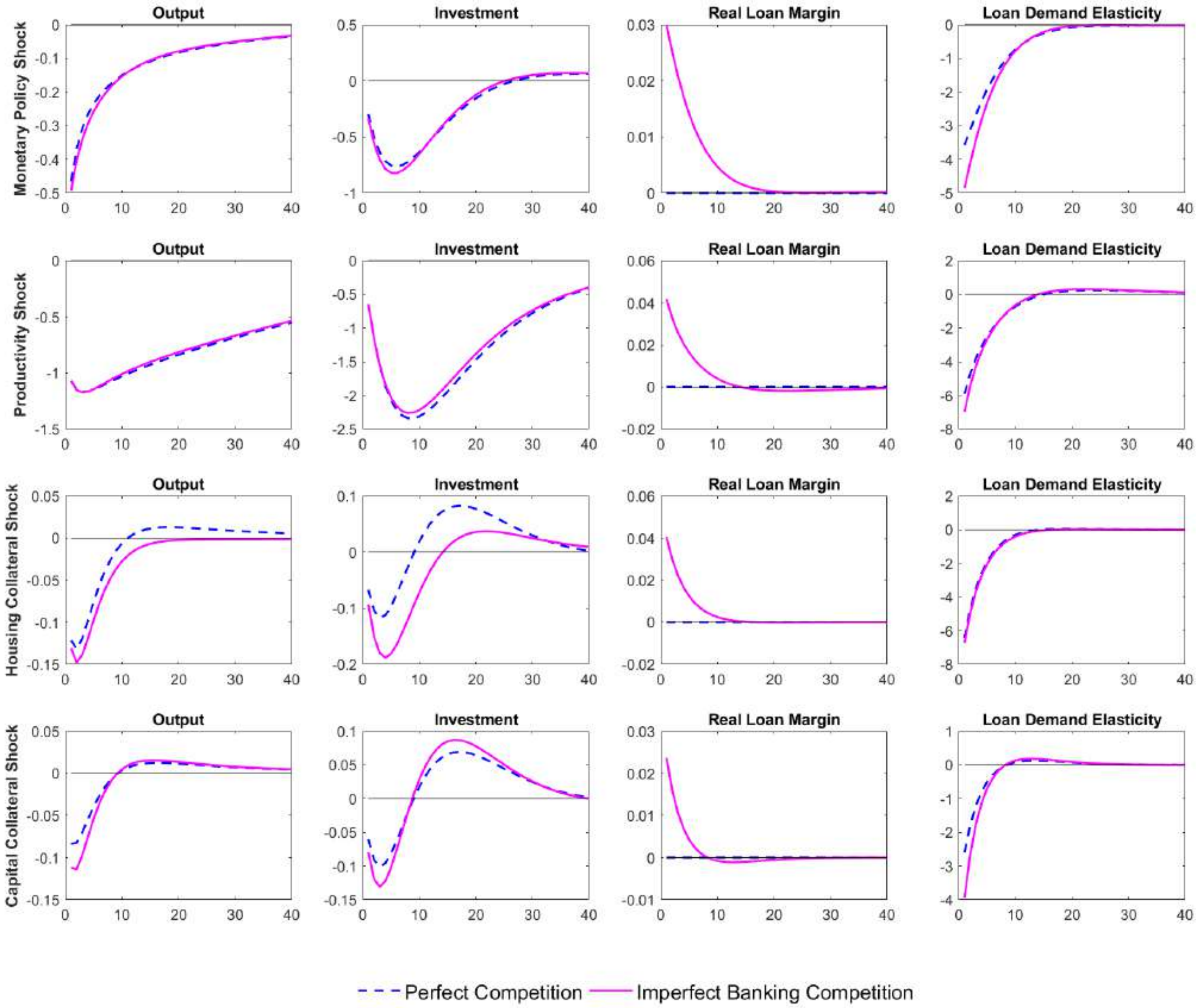
Note: Horizontal axis shows quarters after the shock that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state for variables other than the real loan margin, which is expressed in deviations from the steady state in percent points. The blue dashed line corresponds to perfect banking competition. The pink solid line corresponds to imperfect banking competition. Each row shows the impulse responses of four variables after a given type of shock: a 10 basis-point contractionary monetary policy shock, a one-standard-deviation negative productivity shock, and one-standard-deviation negative shocks to the pledgeability ratios m_h and m_k.

Figure A.3: Impulse Responses for Different Shocks when $\kappa_y = 0.2$



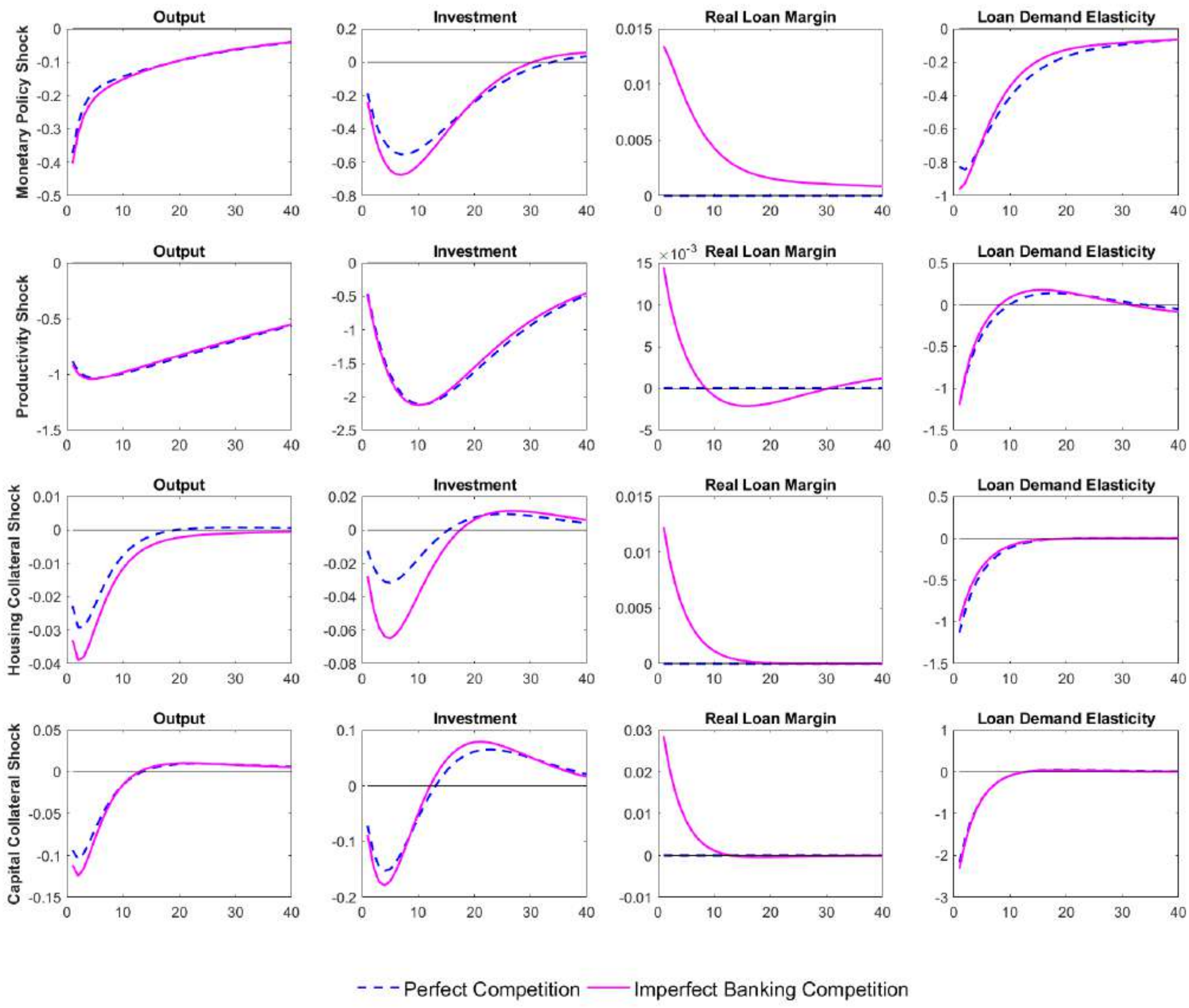
Note: Horizontal axis shows quarters after the shock that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state for variables other than the real loan margin, which is expressed in deviations from the steady state in percent points. The blue dashed line corresponds to perfect banking competition. The pink solid line corresponds to imperfect banking competition. Each row shows the impulse responses of four variables after a given type of shock: a 10 basis-point contractionary monetary policy shock, a one-standard-deviation negative productivity shock, and one-standard-deviation negative shocks to the pledgeability ratios m_h and m_k.

Figure A.4: Impulse Responses for Different Shocks when $m_h = 0.8$ and $m_k = 0.3$



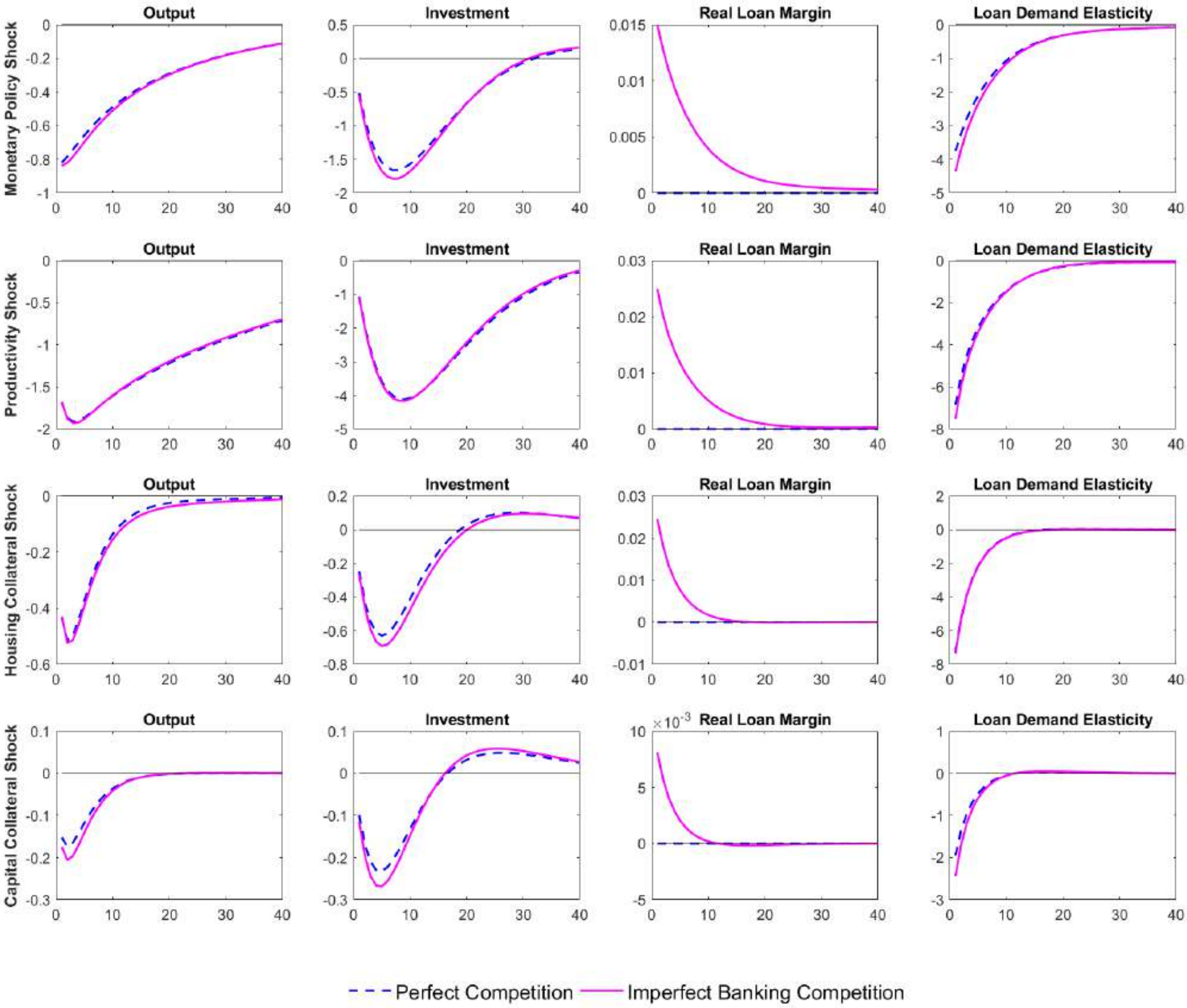
Note: Horizontal axis shows quarters after the shock that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state for variables other than the real loan margin, which is expressed in deviations from the steady state in percent points. The blue dashed line corresponds to perfect banking competition. The pink solid line corresponds to imperfect banking competition. Each row shows the impulse responses of four variables after a given type of shock: a 10 basis-point contractionary monetary policy shock, a one-standard-deviation negative productivity shock, and one-standard-deviation negative shocks to the pledgeability ratios m_h and m_k.

Figure A.5: Impulse Responses for Different Shocks when $m_h = 0.5$ and $m_k = 0.5$



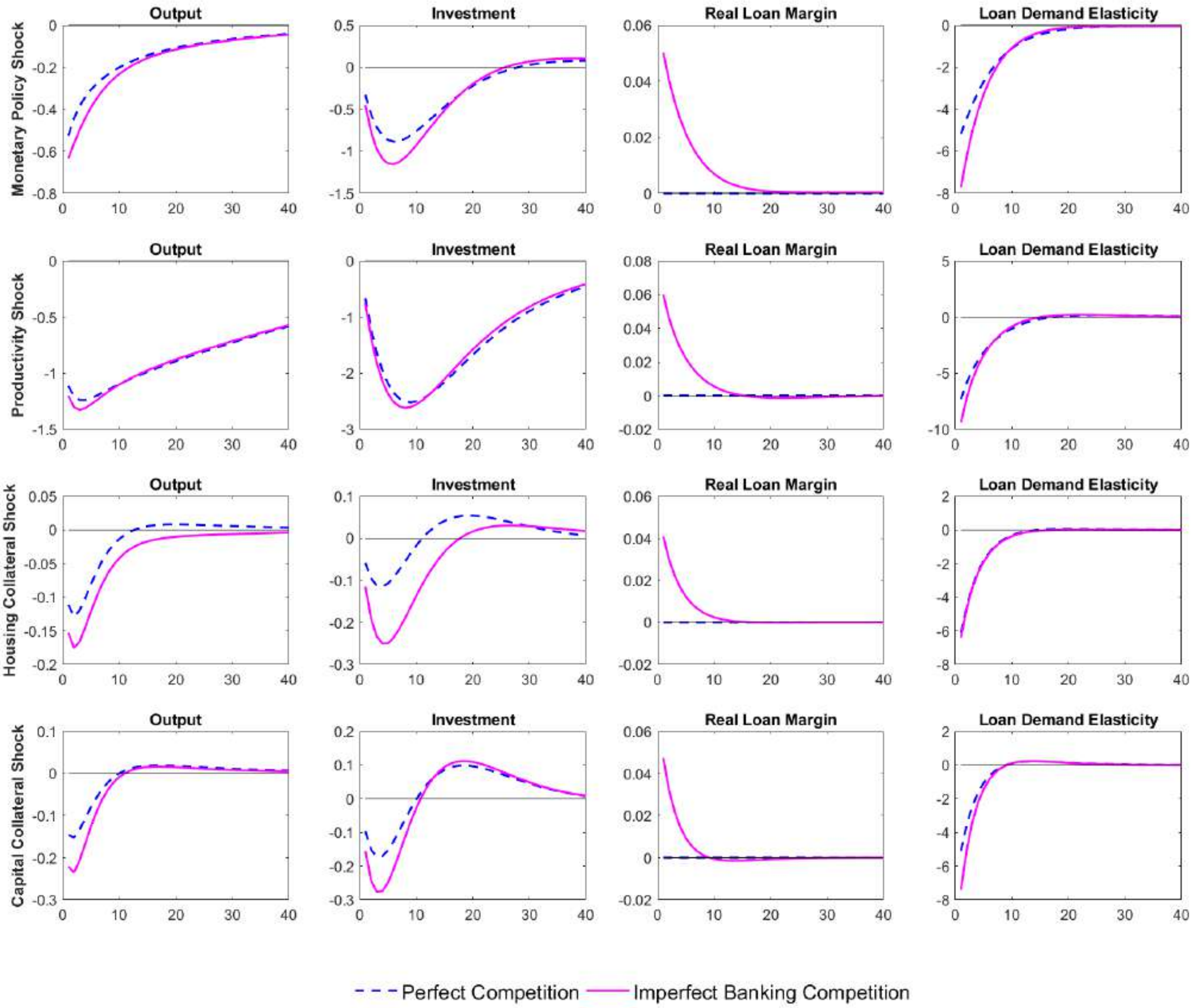
Note: Horizontal axis shows quarters after the shock that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state for variables other than the real loan margin, which is expressed in deviations from the steady state in percent points. The blue dashed line corresponds to perfect banking competition. The pink solid line corresponds to imperfect banking competition. Each row shows the impulse responses of four variables after a given type of shock: a 10 basis-point contractionary monetary policy shock, a one-standard-deviation negative productivity shock, and one-standard-deviation negative shocks to the pledgeability ratios m_h and m_k.

Figure A.6: Impulse Responses for Different Shocks when $\upsilon = 0.15$



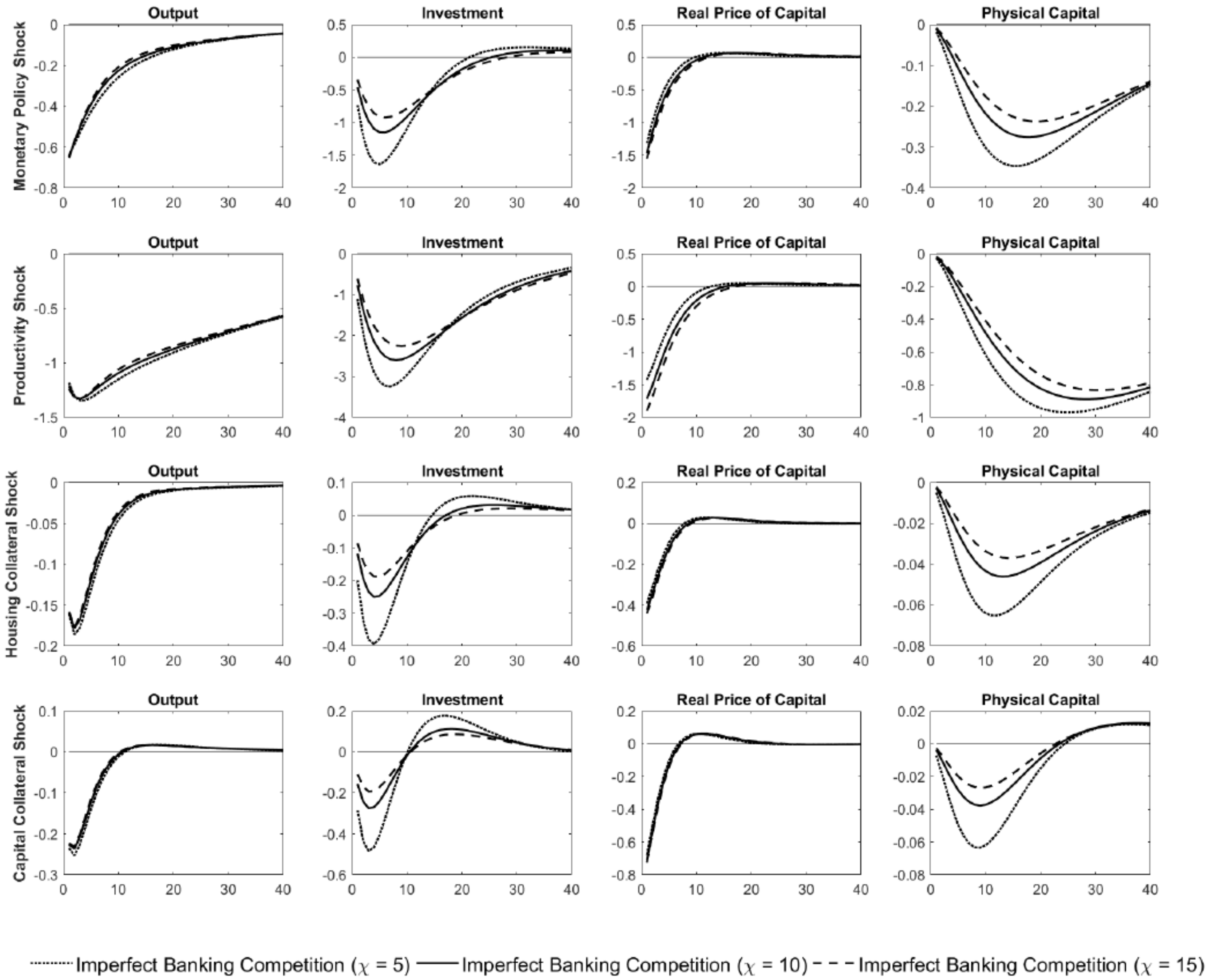
Note: Horizontal axis shows quarters after the shock that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state for variables other than the loan margin, which is expressed in deviations from the steady state in percent points. The blue dashed line corresponds to perfect banking competition. The pink solid line corresponds to imperfect banking competition. Each row shows the impulse responses of four variables after a given type of shock: a 10 basis-point contractionary monetary policy shock, a one-standard-deviation negative productivity shock, and one-standard-deviation negative shocks to the pledgeability ratios m_h and m_k.

Figure A.7: Impulse Responses for Different Shocks when $\phi_h = 0.3$



Note: Horizontal axis shows quarters after the shock that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state for variables other than the real loan margin, which is expressed in deviations from the steady state in percent points. The blue dashed line corresponds to perfect banking competition. The pink solid line corresponds to imperfect banking competition. Each row shows the impulse responses of four variables after a given type of shock: a 10 basis-point contractionary monetary policy shock, a one-standard-deviation negative productivity shock, and one-standard-deviation negative shocks to the pledgeability ratios m_h and m_k.

Figure A.8: Impulse Responses for Different Shocks and Investment Adjustment Cost χ



Note: Horizontal axis shows quarters after the shock that occurs at the beginning of period 1. Vertical axis shows the percentage deviation from the steady state. Each line corresponds to a different calibration for χ under Cournot banking competition ($N = 4$). Each row shows the impulse responses of four variables after a given type of shock: a 10 basis-point contractionary monetary policy shock, a one-standard-deviation negative productivity shock, and one-standard-deviation negative shocks to the pledgeability ratios m_h and m_k.

Appendix B

Appendix to Chapter 3

B.1 Solving the Entrepreneur's Problem

B.1.1 The Slope of the Loan Demand Curve

Rewrite the entrepreneur's expected profit (3.4) as:

$$\begin{aligned} &\mathrm{E}_t\left[\int_{\bar{\omega}_{t+1}(R_{b,t},k_t,\epsilon_{t+1})}^{\infty} \omega\epsilon_{t+1}Ak_t^{\alpha}dF(\omega) - \int_{\bar{\omega}_{t+1}(R_{b,t},k_t,\epsilon_{t+1})}^{\infty} R_{b,t}k_t dF(\omega)\right] \\ =&\mathrm{E}_t\left[\epsilon_{t+1}Ak_t^{\alpha}\int_{\bar{\omega}_{t+1}(R_{b,t},k_t,\epsilon_{t+1})}^{\infty} \omega f(\omega)d\omega - R_{b,t}k_t[1-F(\bar{\omega}_{t+1}(R_{b,t},k_t,\epsilon_{t+1}))]\right] \end{aligned} \tag{B.1}$$

where $f(.)$ is the probability density function (p.d.f.) of the distribution for ω. Recall the condition that determines the entrepreneur's default threshold:

$$\bar{\omega}_{t+1} = \frac{R_{b,t}k_t^{1-\alpha}}{\epsilon_{t+1}A} \tag{3.2}$$

Since $\bar{\omega}_{t+1}$ is a function of k_t, when choosing k_t, the entrepreneur needs to consider the effect of k_t on their default probability $F(\bar{\omega}_{t+1})$. For simplicity, write $\bar{\omega}_{t+1}(R_{b,t}, k_t, \epsilon_{t+1})$ as $\bar{\omega}_{t+1}$ from here onwards. The gross loan rate $R_{b,t}$ is determined by the Cournot banking sector and taken as given by the entrepreneur. Then the first order condition of (B.1) with respect to k_t gives:

$$\begin{aligned} \mathrm{E}_t\Bigg[&\epsilon_{t+1}A\alpha k_t^{\alpha-1}\int_{\bar{\omega}_{t+1}}^{\infty}\omega f(\omega)d\omega - \epsilon_{t+1}Ak_t^{\alpha}\bar{\omega}_{t+1}f(\bar{\omega}_{t+1})\frac{\partial\bar{\omega}_{t+1}}{\partial k_t} \\ &-R_{b,t}[1-F(\bar{\omega}_{t+1})] + R_{b,t}k_t f(\bar{\omega}_{t+1})\frac{\partial\bar{\omega}_{t+1}}{\partial k_t}\Bigg] = 0 \end{aligned} \tag{B.2}$$

Using $\epsilon_{t+1}Ak_t^{\alpha}\bar{\omega}_{t+1} = R_{b,t}k_t$ (3.2), (B.2) can be simplified to:

$$\mathrm{E}_t\left[\epsilon_{t+1}A\alpha k_t^{\alpha-1}\int_{\bar{\omega}_{t+1}}^{\infty}\omega f(\omega)d\omega - R_{b,t}[1-F(\bar{\omega}_{t+1})]\right] = 0 \tag{B.3}$$

Substitute $\epsilon_{t+1}Ak_t^{\alpha-1} = \frac{R_{b,t}}{\bar{\omega}_{t+1}}$ (3.2) into (B.3) and divide each term by $R_{b,t}$ to get:

$$\mathrm{E}_t\left[\frac{\alpha}{\bar{\omega}_{t+1}}\int_{\bar{\omega}_{t+1}}^{\infty}\omega f(\omega)d\omega - [1 - F(\bar{\omega}_{t+1})]\right] = 0 \tag{B.4}$$

The optimal level of k_t for a given $R_{b,t}$ can be solved implicitly from this first order condition. The entrepreneur's default threshold can then be written as $\bar{\omega}_{t+1}(R_{b,t}, k_t(R_{b,t}), \epsilon_{t+1})$.

To check the second order condition, differentiate the LHS of (B.4) with respect to k_t again:

$$\begin{aligned}
&\mathrm{E}_t\left[-\frac{\alpha}{\bar{\omega}_{t+1}^2}\frac{\partial\bar{\omega}_{t+1}}{\partial k_t}\int_{\bar{\omega}_{t+1}}^{\infty}\omega f(\omega)d\omega - \frac{\alpha}{\bar{\omega}_{t+1}}\bar{\omega}_{t+1}f(\bar{\omega}_{t+1})\frac{\partial\bar{\omega}_{t+1}}{\partial k_t} + f(\bar{\omega}_{t+1})\frac{\partial\bar{\omega}_{t+1}}{\partial k_t}\right]\\
=&\mathrm{E}_t\left[-\frac{\alpha}{\bar{\omega}_{t+1}^2}\frac{\partial\bar{\omega}_{t+1}}{\partial k_t}\int_{\bar{\omega}_{t+1}}^{\infty}\omega f(\omega)d\omega + (1-\alpha)f(\bar{\omega}_{t+1})\frac{\partial\bar{\omega}_{t+1}}{\partial k_t}\right]\\
=&\mathrm{E}_t\left[-\frac{\alpha}{\bar{\omega}_{t+1}}(1-\alpha)k_t^{-1}\int_{\bar{\omega}_{t+1}}^{\infty}\omega f(\omega)d\omega + (1-\alpha)^2 f(\bar{\omega}_{t+1})\bar{\omega}_{t+1}k_t^{-1}\right]
\end{aligned} \tag{B.5}$$

where the last step uses $\epsilon_{t+1}\bar{\omega}_{t+1}Ak_t^{\alpha-1} = R_{b,t}$ (3.2), (3.3), and hence $\frac{\partial\bar{\omega}_{t+1}}{\partial k_t} = \frac{(1-\alpha)R_{b,t}k_t^{-\alpha}}{\epsilon_{t+1}A} = (1-\alpha)\bar{\omega}_{t+1}k_t^{-1}$. The second order condition is negative if:

$$\frac{\alpha}{1-\alpha} > \frac{\mathrm{E}_t[f(\bar{\omega}_{t+1})\bar{\omega}_{t+1}]}{\mathrm{E}_t\left[\frac{1}{\bar{\omega}_{t+1}}\int_{\bar{\omega}_{t+1}}^{\infty}\omega f(\omega)d\omega\right]} \tag{B.6}$$

When this condition is satisfied, a unique maximum k_t for a given $R_{b,t}$ can be solved from the first order condition (B.4). Under the calibration in this paper, this condition is always satisfied. Besides, this condition is satisfied if ω has a uniform distribution.

Using the first order condition (B.4) and defining $g(\bar{\omega}_{t+1}) \equiv \frac{\alpha}{\bar{\omega}_{t+1}}\int_{\bar{\omega}_{t+1}}^{\infty}\omega f(\omega)d\omega - [1 - F(\bar{\omega}_{t+1})]$, the slope of the loan demand curve (3.5) can be found using the implicit function theorem:

$$\frac{dk_t}{dR_{b,t}} = -\frac{\mathrm{E}_t\left[\frac{\partial g(\bar{\omega}_{t+1})}{\partial R_{b,t}}\right]}{\mathrm{E}_t\left[\frac{\partial g(\bar{\omega}_{t+1})}{\partial k_t}\right]} = -\frac{\mathrm{E}_t\left[\frac{\partial g(\bar{\omega}_{t+1})}{\partial\bar{\omega}_{t+1}}\frac{\partial\bar{\omega}_{t+1}}{\partial R_{b,t}}\right]}{\mathrm{E}_t\left[\frac{\partial g(\bar{\omega}_{t+1})}{\partial\bar{\omega}_{t+1}}\frac{\partial\bar{\omega}_{t+1}}{\partial k_t}\right]} = -\frac{\mathrm{E}_t\left[\frac{\partial g(\bar{\omega}_{t+1})}{\partial\bar{\omega}_{t+1}}\frac{k_t^{1-\alpha}}{\epsilon_{t+1}A}\right]}{\mathrm{E}_t\left[\frac{\partial g(\bar{\omega}_{t+1})}{\partial\bar{\omega}_{t+1}}\frac{(1-\alpha)k_t^{-\alpha}R_{b,t}}{\epsilon_{t+1}A}\right]} = -\frac{k_t}{(1-\alpha)R_{b,t}} < 0 \tag{3.5}$$

B.1.2 Relationship between the Entrepreneur's Default Threshold and the Gross Loan Rate

Use $\bar{\omega}_{t+1}(R_{b,t}, k_t(R_{b,t}), \epsilon_{t+1})$, where the optimal k_t is a function of $R_{b,t}$, and (3.5) to get (3.6):

$$\frac{d\bar{\omega}_{t+1}}{dR_{b,t}} = \frac{\partial\bar{\omega}_{t+1}}{\partial R_{b,t}} + \frac{\partial\bar{\omega}_{t+1}}{\partial k_t}\frac{dk_t}{dR_{b,t}} = \frac{k_t^{1-\alpha}}{\epsilon_{t+1}A} - \frac{(1-\alpha)k_t^{-\alpha}R_{b,t}}{\epsilon_{t+1}A}\frac{k_t}{(1-\alpha)R_{b,t}} = 0 \tag{3.6}$$

Hence, the gross loan rate does not affect the entrepreneur's default threshold when the entrepreneur is choosing k_t optimally. Alternatively, as can be seen from the total derivative of k_t with respect to $R_{b,t}$ (3.5), a one percent increase in $R_{b,t}$ leads to a $\frac{1}{1-\alpha}$ percent decrease in k_t. Given the expression for the threshold $\bar{\omega}_{t+1} = \frac{R_{b,t}k_t^{1-\alpha}}{\epsilon_{t+1}A}$ (3.2), changes in $R_{b,t}$ will be offset by the endogenous response of k_t, resulting in no overall impact of $R_{b,t}$ on $\bar{\omega}_{t+1}$. In other words, after substituting the optimal k_t for a given level of $R_{b,t}$ into the expression for $\bar{\omega}_{t+1}$, the default threshold $\bar{\omega}_{t+1}(R_{b,t}, k_t(R_{b,t}), \epsilon_{t+1})$ can be simplified to one that only depends on the aggregate shock, i.e., $\bar{\omega}_{t+1}(\epsilon_{t+1})$. This result holds more generally if the entrepreneur is assumed to have full liability, as shown below.

Extension: Entrepreneurs with Full Liability

With full liability, the entrepreneur maximizes the following expected profit with respect to physical capital k_t:

$$\mathrm{E}_t\left[\int_0^\infty \omega\epsilon_{t+1}Ak_t^\alpha dF(\omega) - R_{b,t}k_t\right] = Ak_t^\alpha - R_{b,t}k_t \tag{B.7}$$

where the expectation operator $\mathrm{E}_t[.]$ is taken over the distribution of the aggregate shock ϵ_{t+1} and $\mathrm{E}_t[\epsilon_{t+1}] = 1$. Take the first order condition of (B.7) with respect to k_t:

$$A\alpha k_t^{\alpha-1} - R_{b,t} = 0 \tag{B.8}$$

In this case, the expression of optimal capital demand can be explicitly found from (B.8), which is $k_t = \left(\frac{A\alpha}{R_{b,t}}\right)^{\frac{1}{1-\alpha}}$. The slope of the loan demand curve under full liability is identical to the limited liability case, which can be seen by differentiating the optimal capital demand with respect to $R_{b,t}$. In this case, using the functional form of the default threshold $\bar{\omega}_{t+1}$ (3.2) and the optimal capital demand, the entrepreneur's default threshold can be written as:

$$\bar{\omega}_{t+1} = \frac{\alpha}{\epsilon_{t+1}} \tag{B.9}$$

As can be seen, the entrepreneur's default threshold is still independent of $R_{b,t}$.

To see how the optimal k_t under full liability differs from the one under limited liability, the first order condition under limited liability (B.3) can be rewritten as:

$$A\alpha k_t^{\alpha-1} - R_{b,t} = \mathrm{E}_t\left[\epsilon_{t+1}A\alpha k_t^{\alpha-1}\int_0^{\bar{\omega}_{t+1}} \omega f(\omega)d\omega - R_{b,t}F(\bar{\omega}_{t+1})\right] \tag{B.10}$$

As can be seen, under limited liability of the entrepreneur, the RHS of (B.10) is no longer zero, unlike under full liability of the entrepreneur, when (B.8) holds. Since $A\alpha k_t^{\alpha-1} - R_{b,t}$ decreases in k_t, if the RHS is negative, then k_t under limited liability is larger than its counterpart under full liability.

Simplify the RHS of (B.10) using $\epsilon_{t+1}Ak_t^{\alpha-1} = \frac{R_{b,t}}{\bar{\omega}_{t+1}}$ (3.2):

$$\mathrm{E}_t\left[\frac{\alpha R_{b,t}}{\bar{\omega}_{t+1}}\int_0^{\bar{\omega}_{t+1}}\omega f(\omega)d\omega - R_{b,t}F(\bar{\omega}_{t+1})\right] = \mathrm{E}_t\left[R_{b,t}F(\bar{\omega}_{t+1})\left(\frac{\alpha E[\omega|\omega<\bar{\omega}_{t+1}]}{\bar{\omega}_{t+1}} - 1\right)\right] < 0 \tag{B.11}$$

which is negative as $\frac{E[\omega|\omega<\bar{\omega}_{t+1}]}{\bar{\omega}_{t+1}} < 1$. As a result, k_t under limited liability is larger than its counterpart under full liability. Hence, limited liability leads to a higher $\bar{\omega}_{t+1}$ and thus a higher default probability $F(\bar{\omega}_{t+1})$ than full liability.

B.2 Solving the Bank's Problem

B.2.1 The Equilibrium Gross Loan Rate

Simplify bank j's net profit (3.7) using $\epsilon_{t+1}Ak_t^{\alpha-1} = \frac{R_{b,t}}{\bar{\omega}_{t+1}}$ (3.2) to get:

$$\begin{aligned}
\pi^B_{j,t+1} = &\int_{\bar{\omega}_{t+1}(\epsilon_{t+1})}^{\infty} R_{b,t}k_{j,t}dF(\omega) + \frac{k_{j,t}}{k_t}(1-\mu)\int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})}\epsilon_{t+1}\omega Ak_t^{\alpha}dF(\omega)\\
&- R_t(k_{j,t}-n_{j,t}) - \tau_j k_{j,t} - n_{j,t}\\
=&R_{b,t}k_{j,t}[1-F(\bar{\omega}_{t+1}(\epsilon_{t+1}))] + k_{j,t}(1-\mu)\int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})}\epsilon_{t+1}\omega Ak_t^{\alpha-1}dF(\omega)\\
&- R_t(k_{j,t}-n_{j,t}) - \tau_j k_{j,t} - n_{j,t}\\
=&R_{b,t}k_{j,t}[1-F(\bar{\omega}_{t+1}(\epsilon_{t+1}))] + R_{b,t}k_{j,t}\frac{(1-\mu)}{\bar{\omega}_{t+1}(\epsilon_{t+1})}\int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})}\omega dF(\omega)\\
&- R_t(k_{j,t}-n_{j,t}) - \tau_j k_{j,t} - n_{j,t}\\
=&R_{b,t}k_{j,t}\left[[1-F(\bar{\omega}_{t+1}(\epsilon_{t+1}))] + \frac{(1-\mu)}{\bar{\omega}_{t+1}(\epsilon_{t+1})}\int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})}\omega dF(\omega)\right]\\
&- R_t(k_{j,t}-n_{j,t}) - \tau_j k_{j,t} - n_{j,t}\\
=&R_{b,t}k_{j,t}G(\epsilon_{t+1}) - R_t(k_{j,t}-n_{j,t}) - \tau_j k_{j,t} - n_{j,t}
\end{aligned} \tag{3.8}$$

where

$$\begin{aligned}
G(\epsilon_{t+1}) &\equiv [1-F(\bar{\omega}_{t+1}(\epsilon_{t+1}))] + \frac{1-\mu}{\bar{\omega}_{t+1}(\epsilon_{t+1})}\int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})}\omega f(\omega)d\omega \qquad (B.12)\\
&= [1-F(\bar{\omega}_{t+1}(\epsilon_{t+1}))] + (1-\mu)\frac{E[\omega|\omega\leqslant\bar{\omega}_{t+1}(\epsilon_{t+1})]}{\bar{\omega}_{t+1}(\epsilon_{t+1})}F(\bar{\omega}_{t+1}(\epsilon_{t+1}))\\
&= 1-F(\bar{\omega}_{t+1}(\epsilon_{t+1}))\left[1-(1-\mu)\frac{E[\omega|\omega\leqslant\bar{\omega}_{t+1}(\epsilon_{t+1})]}{\bar{\omega}_{t+1}(\epsilon_{t+1})}\right] < 1
\end{aligned}$$

$G(\epsilon_{t+1}) < 1$ since $\mu \in [0,1]$ and $\frac{E[\omega|\omega<\bar{\omega}_{t+1}(\epsilon_{t+1})]}{\bar{\omega}_{t+1}(\epsilon_{t+1})} < 1$. $G(\epsilon_{t+1})$ denotes the fraction of gross loan return $R_{b,t}k_{j,t}$ that can be obtained by bank j.

$\bar{\omega}_{t+1}$ is a function in terms of only the aggregate shock when the entrepreneur chooses k_t optimally, as shown in (3.6). Due to this result, it is shown below that bank j's choice of loan quantity $k_{j,t}$ does not affect the entrepreneur's default threshold $\bar{\omega}_{t+1}$ in this model, which greatly simplifies the bank's problem. Since the total loan demand k_t is equal to the total loan supply from the j banks, i.e., $k_t = k_{j,t} + \sum_{m \neq j} k_{m,t}$, it follows that under Cournot competition,

$$\frac{dk_t}{dk_{j,t}} = 1 \tag{B.13}$$

Hence, using the fact that $\bar{\omega}_{t+1}$ is independent of $R_{b,t}$ (3.6), the entrepreneur's default threshold is independent of bank j's loan quantity choice $k_{j,t}$ when the entrepreneur is choosing the optimal amount of borrowing:

$$\frac{d\bar{\omega}_{t+1}}{dk_{j,t+1}} = \frac{d\bar{\omega}_{t+1}}{dR_{b,t}} \frac{dR_{b,t}}{dk_t} \frac{dk_t}{dk_{j,t}} = 0 \tag{B.14}$$

A further implication from (B.13) is that the effect of $k_{j,t}$ on the gross loan rate is equivalent to the slope of the downward-sloping inverse demand curve for loans, that is,

$$\frac{dR_{b,t}}{dk_{j,t}} = \frac{dR_{b,t}}{dk_t} \frac{dk_t}{dk_{j,t}} = \frac{dR_{b,t}}{dk_t} \tag{B.15}$$

Using the above three key elements that a) $k_t = k_{j,t} + \sum_{m \neq j} k_{m,t}$, b) $\frac{dR_{b,t}}{dk_{j,t}} = \frac{dR_{b,t}}{dk_t}$ and c) $\frac{d\bar{\omega}_{t+1}}{dk_{j,t}} = 0$, take the first order condition of the expected net profit $E_t[\pi^B_{j,t+1}]$ based on (3.8) with respect to $k_{j,t}$:

$$\left(R_{b,t} + k_{j,t}\frac{dR_{b,t}}{dk_{j,t}}\right) \mathrm{E}_t \left[[1 - F(\bar{\omega}_{t+1}(\epsilon_{t+1}))] + \frac{1-\mu}{\bar{\omega}_{t+1}(\epsilon_{t+1})} \int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})} \omega f(\omega) d\omega \right] - R_t - \tau_j = 0 \tag{B.16}$$

Use (B.15) to replace $\frac{dR_{b,t}}{dk_{j,t}}$ and sum (B.16) over all N banks to get:

$$\begin{aligned} \left(NR_{b,t} + k_t\frac{dR_{b,t}}{dk_t}\right) \mathrm{E}_t \left[[1 - F(\bar{\omega}_{t+1}(\epsilon_{t+1}))] + \frac{1-\mu}{\bar{\omega}_{t+1}(\epsilon_{t+1})} \int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})} \omega f(\omega) d\omega \right] \\ - NR_t - \sum_{j=1}^{N} \tau_j = 0 \end{aligned} \tag{B.17}$$

Since banks have different intermediation costs τ_j, each of them has a different market share in the Cournot equilibrium, depending on their inefficiency indicated by τ_j. Unlike the symmetric case with identical banks where an equilibrium condition $k_{j,t} = \frac{k_t}{N}$ can be imposed, here it is necessary to solve for the equilibrium loan rate and the equilibrium aggregate loan quantity first before knowing the market share of each bank.

Use $\frac{dR_{b,t}}{dk_t} = -\frac{(1-\alpha)R_{b,t}}{k_t}$ from (3.5) and the definition of $G(\epsilon_{t+1})$ in (B.12) to simplify (B.17):

$$R_{b,t}\,(N - 1 + \alpha)\,\mathrm{E}_t\,[G(\epsilon_{t+1})] - NR_t - \sum_{j=1}^{N} \tau_j = 0 \tag{B.18}$$

Rearrange to get the equilibrium gross loan interest rate $R_{b,t}^{*}$ (3.15):

$$R_{b,t}^{*} = \frac{NR_t + \sum_{j=1}^{N} \tau_j}{(N-1+\alpha)\,\mathrm{E}_t\left[G(\epsilon_{t+1})\right]} = \frac{R_t + \bar{\tau}}{\left(1 - \frac{1-\alpha}{N}\right)\mathrm{E}_t[G(\epsilon_{t+1})]} \tag{3.15}$$

where $\bar{\tau} \equiv \frac{1}{N}\sum_{j=1}^{N}\tau_j$ denotes the mean marginal intermediation cost across the N banks. It can be seen that $R_{b,t}^{*} > R_t$ since $\left(1 - \frac{1-\alpha}{N}\right) \leqslant 1$ and $\mathrm{E}_t[G(\epsilon_{t+1})] < 1$.

B.2.2 Parameter Restriction on τ_j

Since τ_j is randomly drawn from an exogenous distribution and the number of banks N is exogenously given, there needs to be a restriction on the value of τ_j to ensure that each of the N banks makes a positive expected profit. More specifically, assume banks are subject to a participation constraint:[1]

$$R_{b,t}k_{j,t}\mathrm{E}_t[G(\epsilon_{t+1})] - R_t(k_{j,t} - n_{j,t}) - \tau_j k_{j,t} > R_t n_{j,t} \tag{B.19}$$

where $\mathrm{E}_t[G(\epsilon_{t+1})] = \mathrm{E}_t\left[[1 - F(\bar{\omega}_{t+1}(\epsilon_{t+1}))] + \frac{1-\mu}{\bar{\omega}_{t+1}(\epsilon_{t+1})}\int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})}\omega f(\omega)d\omega\right]$. The above condition means that bank j with equity $n_{j,t}$ has an incentive to operate only if the profit earned from lending is not less than the opportunity cost of its own funds. Simplify (B.19) to get:

$$R_t + \tau_j < R_{b,t}\mathrm{E}_t[G(\epsilon_{t+1})] \tag{B.20}$$

Substitute the equilibrium loan rate (3.15) to get:

$$R_t + \tau_j < \frac{R_t + \bar{\tau}}{\left(1 - \frac{1-\alpha}{N}\right)} \tag{3.17}$$

which means bank j's marginal cost (the sum of the gross deposit rate and the marginal intermediation cost) cannot be larger than a factor $\frac{1}{\left(1-\frac{1-\alpha}{N}\right)} > 1$ of the mean marginal cost across banks. Note that when banks have identical marginal intermediation cost (i.e., $\tau_j = \bar{\tau}\ \forall j$), the above condition is always satisfied given $\alpha < 1$.

B.2.3 Proof of Proposition 1

Assume the distribution for τ does not change with the number of banks, so the average marginal intermediation cost across banks $\bar{\tau} = \frac{1}{N}\sum_{j=1}^{N}\tau_j$ is an exogenous constant. This is a convenient assumption since the baseline framework focuses on the effect of changing competition (or number of banks) on variables of interest for a given distribution of bank

[1] Banks do not make entry decisions since this paper abstracts away from endogenous entry dynamics. For a given level of N, each operating bank's inefficiency or intermediation cost τ_j is assumed to be within the range that allows each bank to make a positive expected profit.

efficiency. Differentiate the equilibrium loan rate (3.15) with respect to N:

$$\begin{aligned}\frac{dR^*_{b,t}}{dN} &= -\frac{R_t+\bar{\tau}}{(1-\frac{1-\alpha}{N})^2\mathrm{E}_t\left[G(\epsilon_{t+1})\right]}\frac{1-\alpha}{N^2} \\ &= -\frac{(1-\alpha)R^*_{b,t}}{N(N-1+\alpha)} < 0\end{aligned} \tag{B.21}$$

where the second step uses the equilibrium loan rate (3.15). It is straightforward to see that this result is identical to the symmetric case where banks have the same level of efficiency $\tau_j = \bar{\tau}\ \forall j$.

Given the equilibrium loan rate (3.15), the equilibrium total loan quantity k^*_t is also known. It can be shown that k^*_t increases in N:

$$\frac{dk^*_t}{dN} = \frac{dk^*_t}{dR^*_{b,t}}\frac{dR^*_{b,t}}{dN} = -\frac{k^*_t}{(1-\alpha)R^*_{b,t}}\frac{dR^*_{b,t}}{dN} = \frac{k^*_t}{N(N-1+\alpha)} > 0 \tag{B.22}$$

Using (3.1), it can be seen that the expected output is Ak^α_t, as entrepreneurs are ex ante identical. It follows from $\frac{dk^*_t}{dN} > 0$ that the expected output $A(k_t*)^\alpha$ in terms of the optimal k^*_t also increases in N.

Extension: Distribution Mean for τ Changes with N

When the distribution of the marginal intermediation cost τ is allowed to change with N, how the equilibrium loan rate $R^*_{b,t}$ changes with N depends on the efficiency of the new entrants. Using the expression for the equilibrium loan rate (3.15), the change in $R^*_{b,t}$ when N increases by one is:[2]

$$\begin{aligned}R^*_{b,t}(N+1) - R^*_{b,t}(N) =& \left[(N+1)R_t + \sum_{j=1}^{N+1}\tau_j - NR_t - \sum_{j=1}^{N}\tau_j\right]\frac{1}{(N+\alpha)\mathrm{E}_t[G(\epsilon_{t+1})]} \\ &+ \left(NR_t + \sum_{j=1}^{N}\tau_j\right)\left[\frac{1}{(N+\alpha)\mathrm{E}_t[G(\epsilon_{t+1})]} - \frac{1}{(N-1+\alpha)\mathrm{E}_t[G(\epsilon_{t+1})]}\right] \\ =&\frac{(R_t+\tau_{N+1})}{(N+\alpha)\mathrm{E}_t[G(\epsilon_{t+1})]} - \frac{(NR_t+\sum_{j=1}^{N}\tau_j)}{(N+\alpha)(N-1+\alpha)\mathrm{E}_t[G(\epsilon_{t+1})]} \\ =&\frac{(N-1+\alpha)\tau_{N+1} - (1-\alpha)R_t - \sum_{j=1}^{N}\tau_j}{(N+\alpha)(N-1+\alpha)\mathrm{E}_t[G(\epsilon_{t+1})]}\end{aligned} \tag{B.24}$$

where τ_{N+1} denotes the marginal intermediation cost of the new entrant. As can be seen, the sign of $R^*_{b,t}(N+1) - R^*_{b,t}(N)$ depends on the magnitude of the efficiency of the

[2] Using the product rule for discrete functions (sequences) $u(x)$ and $v(x)$, where x denotes the inputs for the discrete functions.

$$\Delta(u(x)v(x)) = \Delta u(x)\Delta v(x) + \Delta u(x)v(x) + u(x)\Delta v(x) = v(x+1)\Delta u(x) + u(x)\Delta v(x) \tag{B.23}$$

where $\Delta u(x) = u(x+1) - u(x)$ and $\Delta v(x) = v(x+1) - v(x)$ are the discrete counterparts of $\frac{du}{dx}$ and $\frac{dv}{dx}$. In this case, let $x = N$, $u(x) = NR_t + \sum_{j=1}^{N}\tau_j$, $v(x) = \frac{1}{(N-1+\alpha)\mathrm{E}_t[G(\epsilon_{t+1})]}$.

$(N+1)$-th bank, τ_{N+1}. This paper focuses on changes in the degree of market power from changes in competition by assuming that the distribution mean for τ is unaffected by N.

B.2.4 Proof of Proposition 2

Once the equilibrium loan rate and the equilibrium aggregate loan quantity are known, each bank j's equilibrium loan quantity $k^*_{j,t}$ (3.18) can be found by using $\frac{dR_{b,t}}{dk_{j,t}} = \frac{dR_{b,t}}{dk_t}$ (B.15), bank j's first order condition (B.16), $\frac{dk^*_t}{dR^*_{b,t}} = -\frac{k^*_t}{(1-\alpha)R^*_{b,t}}$ (3.5), and $R^*_{b,t}$ (3.15):

$$
\begin{aligned}
k^*_{j,t} &= \left[\frac{R_t+\tau_j}{\mathrm{E}_t\left[[1-F(\bar{\omega}_{t+1}(\epsilon_{t+1}))]+\frac{1-\mu}{\bar{\omega}_{t+1}(\epsilon_{t+1})}\int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})}\omega f(\omega)d\omega\right]} - R^*_{b,t}\right]\frac{dk^*_t}{dR^*_{b,t}} \\
&= -\left[\frac{R_t+\tau_j}{\mathrm{E}_t\left[[1-F(\bar{\omega}_{t+1}(\epsilon_{t+1}))]+\frac{1-\mu}{\bar{\omega}_{t+1}(\epsilon_{t+1})}\int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})}\omega f(\omega)d\omega\right]} - R^*_{b,t}\right]\frac{k^*_t}{(1-\alpha)R^*_{b,t}} \\
&= -\frac{k^*_t}{1-\alpha}\left[\frac{R_t+\tau_j}{\mathrm{E}_t\left[[1-F(\bar{\omega}_{t+1}(\epsilon_{t+1}))]+\frac{1-\mu}{\bar{\omega}_{t+1}(\epsilon_{t+1})}\int_0^{\bar{\omega}_{t+1}(\epsilon_{t+1})}\omega f(\omega)d\omega\right]R^*_{b,t}} - 1\right] \\
&= \frac{k^*_t}{1-\alpha}\left[1-\frac{(1-\frac{1-\alpha}{N})(R_t+\tau_j)}{(R_t+\bar{\tau})}\right]
\end{aligned}
\tag{3.18}
$$

Note that in a Cournot equilibrium with heterogeneous banks, bank j's equilibrium market share is no longer equal to $\frac{1}{N}$. According to (3.18), bank j's equilibrium market share $ms^*_{j,t}$ is:

$$
ms^*_{j,t} \equiv \frac{k^*_{j,t}}{k^*_t} = \frac{1}{1-\alpha}\left[1-\frac{(1-\frac{1-\alpha}{N})(R_t+\tau_j)}{(R_t+\bar{\tau})}\right] \tag{B.25}
$$

As can be seen from (B.25), if all banks have the same marginal intermediation cost (i.e., $\tau_j = \tau\ \forall j$), each bank j has a market share of $\frac{1}{N}$. In fact, when the bank has a below average marginal intermediation cost ($\tau_j < \bar{\tau}$), its market share is larger than $\frac{1}{N}$. Given the condition (3.17), bank j's equilibrium market share is positive. Since $ms^*_{j,t} > 0$ and $\sum_{j=1}^{N} ms^*_{j,t} = 1$, each bank's market share is less than 1.

Assume the distribution mean for τ does not change with N, it is shown below that each bank's market share falls with N:

$$
\frac{dms^*_{j,t}}{dN} = -\frac{1}{1-\alpha}\frac{(\frac{1-\alpha}{N^2})(R_t+\tau_j)}{R_t+\bar{\tau}} = -\frac{R_t+\tau_j}{N^2(R_t+\bar{\tau})} < 0 \tag{B.26}
$$

As can be seen, if bank j is more inefficient relative to the average bank (i.e., τ_j is larger than $\bar{\tau}$), then bank j's market share falls by more when N increases. When N is already large, the responsiveness of $ms^*_{j,t}$ to a further increase in N is much smaller.

B.2.5 Proof of Proposition 3

When banks have the same level of efficiency, each bank's loan quantity unambiguously decreases with the number of banks N for $N > 1$. In this case, use $k^*_{j,t} = ms^*_{j,t}k^*_t$, $\frac{dk^*_t}{dN} = \frac{k^*_t}{N(N-1+\alpha)}$ (B.22) and $ms^*_{j,t} = \frac{1}{N}$ to get:

$$\begin{aligned}\frac{dk^*_{j,t}}{dN} =& ms^*_{j,t}\frac{dk^*_t}{dN} + \frac{dms^*_{j,t}}{dN}k^*_t \\ =& \frac{1}{N}\frac{k^*_t}{N(N-1+\alpha)} - \frac{1}{N^2}k^*_t \\ =& \left(\frac{1}{N-1+\alpha} - 1\right)\frac{k^*_t}{N^2} < 0 \quad \text{if } N > 1\end{aligned} \tag{B.27}$$

By contrast, when banks have different levels of efficiency, how each individual bank's loan quantity changes with N is unclear, depending on the balance between an increasing aggregate loan quantity k^*_t and the falling equilibrium market share when N increases (Proposition 2). Using $k^*_{j,t} = ms^*_{j,t}k^*_t$, the expressions for $ms^*_{j,t}$ (B.25), $\frac{dk^*_t}{dN}$ (B.22) and $\frac{dms_{j,t}}{dN}^*$ (B.26), it is shown below that the sign of $\frac{dk^*_{j,t}}{dN}$ is ambiguous:

$$\begin{aligned}\frac{dk^*_{j,t}}{dN} =& ms^*_{j,t}\frac{dk^*_t}{dN} + \frac{dms^*_{j,t}}{dN}k^*_t \\ =& \frac{k^*_{j,t}}{k^*_t}\frac{k^*_t}{N(N-1+\alpha)} - \frac{R_t+\tau_j}{N^2(R_t+\bar{\tau})}k^*_t \\ =& \frac{1}{1-\alpha}\left[1 - \frac{(1-\frac{1-\alpha}{N})(R_t+\tau_j)}{(R_t+\bar{\tau})}\right]\frac{k^*_t}{N(N-1+\alpha)} - \frac{R_t+\tau_j}{N^2(R_t+\bar{\tau})}k^*_t \\ =& \frac{[R_t+\bar{\tau} - (1-\frac{1-\alpha}{N})(R_t+\tau_j)]k^*_t - \frac{N-1+\alpha}{N}(1-\alpha)(R_t+\tau_j)k^*_t}{(1-\alpha)(R_t+\bar{\tau})N(N-1+\alpha)} \\ =& \frac{[R_t+\bar{\tau} - (2-\alpha)(1-\frac{1-\alpha}{N})(R_t+\tau_j)]k^*_t}{(1-\alpha)(R_t+\bar{\tau})N(N-1+\alpha)}\end{aligned} \tag{B.28}$$

It follows from (B.28) that $\frac{dk^*_{j,t}}{dN} < 0$ when the numerator is negative:

$$R_t + \bar{\tau} - (2-\alpha)\left(1 - \frac{1-\alpha}{N}\right)(R_t+\tau_j) < 0 \tag{B.29}$$

or equivalently after rearranging,

$$R_t + \tau_j > \frac{R_t+\bar{\tau}}{(2-\alpha)(1-\frac{1-\alpha}{N})} \tag{B.30}$$

Since $\frac{R_t+\bar{\tau}}{(2-\alpha)(1-\frac{1-\alpha}{N})}$ is strictly smaller than the upper bound $\frac{R_t+\bar{\tau}}{(1-\frac{1-\alpha}{N})}$ (3.17), there is a positive probability that $\frac{dk^*_{j,t}}{dN}$ is negative for some banks and positive for others depending on the bank's relative efficiency. More specifically, $\frac{dk^*_{j,t}}{dN} < 0$ when

$$\frac{R_t+\bar{\tau}}{(2-\alpha)(1-\frac{1-\alpha}{N})} < R_t + \tau_j < \frac{R_t+\bar{\tau}}{(1-\frac{1-\alpha}{N})} \tag{B.31}$$

and $\frac{dk^*_{j,t}}{dN} > 0$ when

$$R_t + \tau_j < \frac{R_t + \bar{\tau}}{(2-\alpha)(1 - \frac{1-\alpha}{N})} \tag{B.32}$$

Intuitively, although each bank's market share falls with N (Proposition 2), this effect of market share reduction can be offset by the increase in total loan quantity as N increases, leading to an increase in bank j's loan quantity. According to (B.26), the market shares of more efficient banks with low τ_j relative to the mean are less sensitive to changes in N. So an increase in aggregate loan quantity as N increases can be large relative to a small drop in market share of a more efficient bank, resulting in an increase in the bank's loan quantity.

B.2.6 Proof of Proposition 4

It can be shown that the effect of changes in $k^*_{j,t}$ in response to an increase in N is dominated by the effect of the fall in $R^*_{b,t}$, so the expected net profit $\mathrm{E}_t[\pi^B_{t+1}]$ decreases with N. Following (3.8), the expected net profit in equilibrium is:

$$\mathrm{E}_t[\pi^B_{j,t+1}] \equiv R^*_{b,t} k^*_{j,t} \mathrm{E}_t[G(\epsilon_{t+1})] - R_t(k^*_{j,t} - n_{j,t}) - \tau_j k^*_{j,t} - n_{j,t} \tag{B.33}$$

Differentiate $\mathrm{E}_t[\pi^B_{j,t+1}]$ with respect to N and use the expressions for $\frac{dR^*_{b,t}}{dN}$ (B.21) and $\frac{dk^*_{j,t}}{dN}$ (B.28) to get:

$$\begin{aligned}
\frac{d\mathrm{E}_t[\pi^B_{j,t+1}]}{dN} =& \frac{dR^*_{b,t}}{dN} k^*_{j,t} \mathrm{E}_t[G(\epsilon_{t+1})] + \frac{dk^*_{j,t}}{dN} \left(R^*_{b,t} \mathrm{E}_t[G(\epsilon_{t+1})] - R_t - \tau_j \right) \\
=& - \frac{(1-\alpha) R^*_{b,t} k^*_{j,t} \mathrm{E}_t[G(\epsilon_{t+1})]}{N(N-1+\alpha)} \\
&+ \left(\frac{k^*_{j,t}}{N(N-1+\alpha)} - \frac{(R_t+\tau_j)}{N^2(R_t+\bar{\tau})} k^*_t \right) \left(R^*_{b,t} \mathrm{E}_t[G(\epsilon_{t+1})] - R_t - \tau_j \right) \\
=& \frac{\alpha R^*_{b,t} k^*_{j,t} \mathrm{E}_t[G(\epsilon_{t+1})] - (R_t+\tau_j) k^*_{j,t}}{N(N-1+\alpha)} - \frac{(R_t+\tau_j)}{N^2(R_t+\bar{\tau})} k^*_t \left(R^*_{b,t} \mathrm{E}_t[G(\epsilon_{t+1})] - R_t - \tau_j \right) < 0
\end{aligned} \tag{B.34}$$

Proof for $\frac{d\mathrm{E}_t[\pi^B_{t+1}]}{dN} < 0$:

1) According to (B.20):

$$R^*_{b,t} \mathrm{E}_t[G(\epsilon_{t+1})] - R_t - \tau_j > 0 \tag{B.20}$$

2) Using $ms^*_{j,t}$ (B.25), the fact that $ms^*_{j,t} < 1$ gives:

$$\frac{(1 - \frac{1-\alpha}{N})(R_t + \tau_j)}{(R_t + \bar{\tau})} > \alpha \tag{B.35}$$

Substitute $R^*_{b,t} \mathrm{E}_t[G(\epsilon_{t+1})] = \frac{R_t + \bar{\tau}}{1 - \frac{1-\alpha}{N}}$ (3.15) into the above inequality and rearrange to get:

$$\alpha R^*_{b,t} k^*_{j,t} \mathrm{E}_t[G(\epsilon_{t+1})] - (R_t + \tau_j) k^*_{j,t} < 0 \tag{B.36}$$

B.2.7 Proof of Proposition 5

Given the predetermined equity $n_{j,t}$, bank j chooses the loan quantity $k_{j,t}$ to maximize $E_t[\pi^B_{j,t+1}]$. In the presence of an adverse aggregate shock in period $t+1$, the net profit $\pi^B_{j,t+1}$ can be negative and if the loss is too large to be absorbed by the equity $n_{j,t}$, the pre-dividend equity $n_{j,t}+\pi^B_{j,t+1}$ in period $t+1$ would be negative in which case bank j defaults. More specifically, if the realized value of the aggregate shock ϵ_{t+1} is below bank j's default threshold $\bar{\epsilon}_{j,t+1}$, then bank j becomes insolvent, where $\bar{\epsilon}_{j,t+1}$ is determined by the condition (3.21):

$$n_{j,t}+\pi^B_{j,t+1}(\bar{\epsilon}_{j,t+1}) = R^*_{b,t}k^*_{j,t}G(\bar{\epsilon}_{j,t+1}) - R_t(k^*_{j,t}-n_{j,t}) - \tau_j k^*_{j,t} = 0 \tag{3.21}$$

where $G(\bar{\epsilon}_{j,t+1}) \equiv \Big[[1-F(\bar{\omega}_{t+1}(\bar{\epsilon}_{j,t+1}))] + \frac{1-\mu}{\bar{\omega}_{t+1}(\bar{\epsilon}_{j,t+1})}\int_0^{\bar{\omega}_{t+1}(\bar{\epsilon}_{j,t+1})}\omega f(\omega)d\omega\Big] < 1$. $G(\bar{\epsilon}_{j,t+1})$ is a fraction of the contractual gross loan revenue $R_{b,t}k_{j,t}$ that can be earned by bank j when the realized aggregate shock takes a value of $\bar{\epsilon}_{j,t+1}$. This condition shows that the pre-dividend equity in period $t+1$ is zero when the realized value of the aggregate shock is $\bar{\epsilon}_{j,t+1}$. It can be shown that $G'(\bar{\epsilon}_{j,t+1})$ is positive:

$$\begin{aligned}
G'(\bar{\epsilon}_{j,t+1}) &= -f(\bar{\omega}_{t+1})\frac{\partial\bar{\omega}_{t+1}}{\partial\bar{\epsilon}_{j,t+1}} - \frac{1-\mu}{\bar{\omega}^2_{t+1}}\frac{\partial\bar{\omega}_{t+1}}{\partial\bar{\epsilon}_{j,t+1}}\int_0^{\bar{\omega}_{t+1}}\omega f(\omega)d\omega + \frac{1-\mu}{\bar{\omega}_{t+1}}\bar{\omega}_{t+1}f(\bar{\omega}_{t+1})\frac{\partial\bar{\omega}_{t+1}}{\partial\bar{\epsilon}_{j,t+1}} \\
&= \frac{\partial\bar{\omega}_{t+1}}{\partial\bar{\epsilon}_{j,t+1}}\left[-f(\bar{\omega}_{t+1}) - \frac{1-\mu}{\bar{\omega}^2_{t+1}}\int_0^{\bar{\omega}_{t+1}}\omega f(\omega)d\omega + (1-\mu)f(\bar{\omega}_{t+1})\right] \\
&= -\frac{\partial\bar{\omega}_{t+1}}{\partial\bar{\epsilon}_{j,t+1}}\left[\frac{1-\mu}{\bar{\omega}^2_{t+1}}\int_0^{\bar{\omega}_{t+1}}\omega f(\omega)d\omega + \mu f(\bar{\omega}_{t+1})\right] > 0
\end{aligned} \tag{B.37}$$

where $\frac{\partial\bar{\omega}_{t+1}}{\partial\bar{\epsilon}_{j,t+1}} < 0$, as can be seen from the entrepreneur's ex post default threshold $\bar{\omega}_{t+1}(\bar{\epsilon}_{j,t+1}) = \frac{R_{b,t}k^{*1-\alpha}_{j,t}}{\bar{\epsilon}_{j,t+1}A}$ based on (3.2), when the realized aggregate shock is $\bar{\epsilon}_{j,t+1}$. Intuitively, a higher realized aggregate shock reduces the entrepreneur's default threshold and thus default probability. Hence, $G(.)$ increases in $\bar{\epsilon}_{j,t+1}$, implying higher realized aggregate shock raises the fraction of $R^*_{b,t}k^*_{j,t}$ that can be obtained by bank j. Bank's default condition (3.21) shows that when the realized aggregate shock is below $\bar{\epsilon}_{j,t+1}$, the loan revenue $R^*_{b,t}k^*_{j,t}G(\bar{\epsilon}_{j,t+1})$ is too low and hence the net profit is too negative to be absorbed by $n_{j,t}$ such that bank j has to default.

Banks have different default thresholds due to different marginal intermediation costs τ_j and the predetermined equity $n_{j,t}$. Divide each term in (3.21) by $k_{j,t}$ to get (3.22):

$$R^*_{b,t}G(\bar{\epsilon}_{j,t+1}) - (R_t+\tau_j) + R_t\frac{n_{j,t}}{k^*_{j,t}} = 0 \tag{3.22}$$

Implicitly differentiate (3.22) with respect to the equity ratio $\kappa_{j,t} = \frac{n_{j,t}}{k^*_{j,t}}$ to get:

$$R^*_{b,t}G'(\bar{\epsilon}_{j,t+1})\frac{d\bar{\epsilon}_{j,t+1}}{d\kappa_{j,t}} + R_t = 0 \tag{B.38}$$

Rearrange to get:

$$\frac{d\bar{\epsilon}_{j,t+1}}{d\kappa_{j,t}} = -\frac{R_t}{R^*_{b,t}G'(\bar{\epsilon}_{j,t+1})} < 0 \tag{B.39}$$

B.2.8 Proof of Proposition 6

As can be seen from (3.22), a change in the number of banks N can affect bank j's default threshold via the profit margin $[R_{b,t}G(\bar{\epsilon}_{j,t+1}) - (R_t + \tau_j)]$, which resembles the margin effect. A lower N raises $R_{b,t}$ and hence the profit margin for an exogenous marginal cost $(R_t + \tau_j)$, resulting in a lower default threshold (margin effect). The equity ratio $\frac{n_{j,t}}{k^*_{j,t}}$ present in (3.22) reflects the equity ratio effect. A higher equity ratio lowers the default threshold since the bank is still able to survive with a lower realized aggregate shock. A lower N can lead to a larger $k_{j,t}$ (Proposition 3) and thus a lower equity ratio $\frac{n_{j,t}}{k_{j,t}}$ since $n_{j,t}$ is predetermined. This short-run equity ratio effect tends to raise the bank's default threshold as N decreases, which opposes the margin effect. Note that if $n_{j,t} = 0$, the short-run equity ratio effect is absent and an increase in N unambiguously raises bank j's default threshold. A formal proof is shown below. Totally differentiate (3.22) with respect to N:

$$R^*_{b,t}G'(\bar{\epsilon}_{j,t+1})\frac{d\bar{\epsilon}_{j,t+1}}{dN} + \frac{dR^*_{b,t}}{dN}G(\bar{\epsilon}_{j,t+1}) + R_t\frac{1}{k^*_{j,t}}\frac{dn_{j,t}}{dN} - R_t\frac{n_{j,t}}{(k^*_{j,t})^2}\frac{dk^*_{j,t}}{dN} = 0 \tag{B.40}$$

Rearrange to get (3.23):

$$\frac{d\bar{\epsilon}_{j,t+1}}{dN} = \frac{R_t\frac{n_{j,t}}{k^*_{j,t}}\frac{dk^*_{j,t}}{dN}\frac{1}{k^*_{j,t}} - \frac{dR^*_{b,t}}{dN}G(\bar{\epsilon}_{j,t+1}) - R_t\frac{1}{k^*_{j,t}}\frac{dn_{j,t}}{dN}}{R^*_{b,t}G'(\bar{\epsilon}_{j,t+1})} \tag{3.23}$$

Since equity in period t is predetermined and is not affected by changes in N in period t, $\frac{dn_{j,t}}{dN} = 0$. However, future equity levels will be affected by changes in N, so do future default probabilities. Hence, the last term in the numerator refers to the long-run equity ratio effect.

As can be seen from (3.23), when $n_{j,t} = 0$, the sign of $\frac{d\bar{\epsilon}_{j,t+1}}{dN}$ is unambiguously positive due to the margin effect. When $n_{j,t} \neq 0$, the sign of $\frac{d\bar{\epsilon}_{j,t+1}}{dN}$ is ambigous, as proved below.

Use the expression for $\frac{dk^*_{j,t}}{dN_t}$ (B.28) and $k^*_{j,t}$ (3.18) to get:

$$\begin{aligned}\frac{dk^*_{j,t}}{dN}\frac{1}{k^*_{j,t}} &= \frac{[R_t + \bar{\tau} - (2-\alpha)(1-\frac{1-\alpha}{N})(R_t+\tau_j)]k^*_t}{(1-\alpha)(R_t+\bar{\tau})N(N-1+\alpha)}\frac{1-\alpha}{k^*_t}\frac{R_t+\bar{\tau}}{R_t+\bar{\tau}-(1-\frac{1-\alpha}{N})(R_t+\tau_j)} \\ &= \frac{R_t + \bar{\tau} - (2-\alpha)(1-\frac{1-\alpha}{N})(R_t+\tau_j)}{N(N-1+\alpha)[R_t+\bar{\tau}-(1-\frac{1-\alpha}{N})(R_t+\tau_j)]}\end{aligned} \tag{B.41}$$

Substitute (B.41) and the expression for $\frac{dR^*_{b,t}}{dN}$ (B.21) into (3.23):

$$
\begin{aligned}
\frac{d\bar{\epsilon}_{j,t+1}}{dN} &= \frac{R_t \frac{n_{j,t}}{k^*_{j,t}} \frac{R_t+\bar{\tau}-(2-\alpha)(1-\frac{1-\alpha}{N})(R_t+\tau_j)}{N(N-1+\alpha)[R_t+\bar{\tau}-(1-\frac{1-\alpha}{N})(R_t+\tau_j)]} + \frac{(1-\alpha)R^*_{b,t}}{N(N-1+\alpha)} G(\bar{\epsilon}_{j,t+1})}{R^*_{b,t} G'(\bar{\epsilon}_{j,t+1})} \\
&= \frac{R_t \frac{n_{j,t}}{k^*_{j,t}} \frac{R_t+\bar{\tau}-(2-\alpha)(1-\frac{1-\alpha}{N})(R_t+\tau_j)}{N(N-1+\alpha)[R_t+\bar{\tau}-(1-\frac{1-\alpha}{N})(R_t+\tau_j)]} + \frac{(1-\alpha)}{N(N-1+\alpha)} \left[R_t(1-\frac{n_{j,t}}{k^*_{j,t}}) + \tau_j\right]}{R^*_{b,t} G'(\bar{\epsilon}_{j,t+1})} \\
&= \frac{R_t \frac{n_{j,t}}{k^*_{j,t}} \frac{1}{N(N-1+\alpha)} \left[\frac{R_t+\bar{\tau}-(2-\alpha)(1-\frac{1-\alpha}{N})(R_t+\tau_j)}{[R_t+\bar{\tau}-(1-\frac{1-\alpha}{N})(R_t+\tau_j)]} - (1-\alpha)\right] + \frac{(1-\alpha)}{N(N-1+\alpha)}(R_t+\tau_j)}{R^*_{b,t} G'(\bar{\epsilon}_{j,t+1})} \\
&= \frac{R_t \frac{n_{j,t}}{k^*_{j,t}} \frac{\alpha(R_t+\bar{\tau})-(1-\frac{1-\alpha}{N})(R_t+\tau_j)}{R_t+\bar{\tau}-(1-\frac{1-\alpha}{N})(R_t+\tau_j)} + (1-\alpha)(R_t+\tau_j)}{N(N-1+\alpha) R^*_{b,t} G'(\bar{\epsilon}_{j,t+1})}
\end{aligned}
\tag{B.42}
$$

where the second step uses (3.22). Since $G'(\bar{\epsilon}_{j,t+1}) > 0$ (B.37), $\frac{d\bar{\epsilon}_{j,t+1}}{dN}$ (B.42) is negative (short-run equity ratio effect dominates the margin effect) if the numerator of (B.42) is negative, or equivalently,

$$
\frac{n_{j,t}}{k^*_{j,t}} > \frac{(1-\alpha)(R_t+\tau_j)[R_t+\bar{\tau}-(1-\frac{1-\alpha}{N})(R_t+\tau_j)]}{R_t[(1-\frac{1-\alpha}{N})(R_t+\tau_j)-\alpha(R_t+\bar{\tau})]} > 0 \tag{B.43}
$$

where $R_t + \bar{\tau} - (1 - \frac{1-\alpha}{N})(R_t + \tau_j) > 0$ (3.17) and $(1 - \frac{1-\alpha}{N})(R_t + \tau_j) - \alpha(R_t + \bar{\tau}) > 0$ (B.35). Rearrange (3.19) to get:

$$
\alpha(R_t+\bar{\tau}) < \left(1-\frac{1-\alpha}{N}\right)(R_t+\tau_j) < R_t+\bar{\tau} \tag{3.19}
$$

So the ratio $\frac{R_t+\bar{\tau}-(1-\frac{1-\alpha}{N})(R_t+\tau_j)}{(1-\frac{1-\alpha}{N})(R_t+\tau_j)-\alpha(R_t+\bar{\tau})}$ on the right hand side of the inequality (B.43) can be larger or smaller than one depending on the value of τ_j. If τ_j is relatively large, the ratio is smaller and it is more likely for the inequality (B.43) to hold. This means when N is lower, the default thresholds of relatively inefficient banks are more likely to increase due to a stronger short-run equity ratio effect ($k_{j,t}$ increases more after a decrease in N) and a weaker margin effect (profit margin is smaller due to higher τ_j).

B.3 Simulation

B.3.1 Reverse Bounded Pareto Distribution for τ

Suppose τ has a Pareto distribution, then the p.d.f. $f_\tau(\tau)$ and c.d.f. $F_\tau(\tau)$ are:

$$
f_\tau(\tau) = \frac{a\tau_s^a}{\tau^{a+1}} \tag{B.44}
$$

$$
F_\tau(\tau) = 1 - \left(\frac{\tau_s}{\tau}\right)^a \tag{B.45}
$$

where $\tau_s > 0$ is the scale parameter and $a > 0$ is the shape parameter and the support is $\tau \in [\tau_s, \infty)$. Bounded (truncated) Pareto distribution is a conditional distribution that results from restricting the domain of Pareto distribution. By restricting the domain of the Pareto distribution to $(L, H]$, the p.d.f. $f_{\tau B}(\tau)$ and c.d.f. $F_{\tau B}(\tau)$ of the bounded Pareto distribution are respectively:

$$f_{\tau B}(\tau) = \frac{f_\tau(\tau)}{F_\tau(H) - F_\tau(L)} = \frac{\frac{a\tau_s^a}{\tau^{a+1}}}{1 - \left(\frac{\tau_s}{H}\right)^a - [1 - \left(\frac{\tau_s}{L}\right)^a]} = \frac{aL^a\tau^{-a-1}}{1 - \left(\frac{L}{H}\right)^a} \tag{B.46}$$

$$F_{\tau B}(\tau) = \frac{F_\tau(\tau) - F_\tau(L)}{F_\tau(H) - F_\tau(L)} = \frac{1 - \left(\frac{\tau_s}{\tau}\right)^a - [1 - \left(\frac{\tau_s}{L}\right)^a]}{1 - \left(\frac{\tau_s}{H}\right)^a - [1 - \left(\frac{\tau_s}{L}\right)^a]} = \frac{1 - L^a\tau^{-a}}{1 - \left(\frac{L}{H}\right)^a} \tag{B.47}$$

where the support is $\tau \in (L, H]$. The bounded Pareto distribution is positively skewed with a long right tail in the domain of $(L, H]$. To generate a market share distribution that contains a few large banks and a lot of small banks, this distribution for τ needs to be reversed such that it is negatively skewed with a long left tail since small τ implies large equilibrium market share. So the p.d.f. of the bounded Pareto distribution is flipped around the y-axis and then shifted to the right by $L + H$, leading to a reverse bounded Pareto distribution that lies within the same domain $(L, H]$. Using (B.46), the p.d.f. of the reverse distribution $f_{\tau BR}(\tau)$ becomes:

$$f_{\tau BR}(\tau) \equiv f_{\tau B}(-\tau + H + L) = \frac{aL^a(H + L - \tau)^{-a-1}}{1 - \left(\frac{L}{H}\right)^a} \tag{B.48}$$

Hence, the c.d.f. of the reverse distribution $F_{\tau BR}(\tau)$ is:

$$F_{\tau BR}(\tau) = \int_L^\tau \frac{aL^a(H + L - \tau)^{-a-1}}{1 - \left(\frac{L}{H}\right)^a} d\tau = \frac{L^a(H + L - \tau)^{-a} - L^aH^{-a}}{1 - \left(\frac{L}{H}\right)^a} \tag{B.49}$$

τ_j is drawn from the reverse bounded Pareto distribution $F_{\tau BR}(\tau)$ with domain $(L, H]$. Applying the inverse-transform method, this distribution can be generated using a uniform distribution $Uniform[0, 1]$. Let U denote a random variable with the continuous uniform distribution over the interval [0, 1], τ_j can be drawn from $F^{-1}_{\tau BR}(U)$, where $F^{-1}_{\tau BR}(.)$ represents the inverse function. The inverse transform method can be used as long as there is an explicit expression for $F^{-1}_{\tau BR}(.)$ in closed form. Using the expression for (B.49),

$$U = \frac{L^a(H + L - \tau)^{-a} - L^aH^{-a}}{1 - \left(\frac{L}{H}\right)^a} \tag{B.50}$$

Rearrange the above equation for τ:

$$\tau = H + L - [UL^{-a} - UH^{-a} + H^{-a}]^{-\frac{1}{a}} \tag{B.51}$$

In simulation, random numbers are first generated from a uniform distribution $U[0, 1]$, then τ_j is obtained using (B.51).

B.3.2 Calibration

Table B.1: Baseline Calibration of Parameters

Parameter	Value Germany
Number of banks N	60
Capital share α	0.3
Desired equity ratio κ^*	0.072
Collection cost μ	0.04
Support for bounded Pareto distribution of τ	[0.001, 0.04]
Shape for bounded Pareto distribution of τ	0.1
Mean of log-normal distribution of ω	-0.15
Variance of log-normal distribution of ω	0.3
Mean of log-normal distribution of ϵ	-0.14
Variance of log-normal distribution of ϵ	0.28

B.4 Data

B.4.1 Data Cleaning

Credit Default Swaps from the Thomson Reuters

Banks with 5-year CDS traded are identified by their names or Ticker in EIKON database. From the download for all 5-year CDS data at a quarterly frequency, there are 306 banks from all countries and 218 unique banks in EU or OECD countries. Each bank can have multiple CDS securities, with different seniorities, currencies, restructuring events, or data providers, which are uniquely identified by RIC (Reuters instrument code) in EIKON database.[3] There are 4103 unique RIC (CDS securities) from all countries and 3534 unique RIC for banks in EU or OECD countries from the download.

After dropping the missing CDS midspread data (302 banks left), the following steps are taken to make sure only one CDS security is kept for each bank:

1) Keep only one type of seniority for each bank. There are 4 types of issue seniority: junior, secured, senior unsecured (67%) and subordinated (≈33%). The last two types

[3]Restructuring event is one type of credit events that triggers settlement under the CDS contract. Restructuring event is a "soft event" as the loss to the owner of the specific bond referenced in the CDS contract is not obvious.

account for the majority of the data points. For each bank, keep the seniority type that occurs most frequently throughout time to maximize the number of data points. After this step, only two types of seniority are left: senior unsecured (≈ 96%) and subordinated (≈ 4%), and 300 banks are left.

2) For each bank, keep the restructuring event that appears most frequently throughout time. After this step, 296 banks are left.

3) Keep only one type of currency for each bank. There are 18 different currencies, with Euro and US dollar accounting for approximately 46% and 40% of the data points respectively. For each bank, keep the currency that occurs most frequently throughout time. After this step, 4 types of currencies (Australian Dollar, British Pound Sterling, Euro, Japanese Yen, and US Dollar) and 292 banks are left.

4) Keep only one data contributor for each bank. There are 12 different data contributors. GFI FENICS (≈ 36%), Thomson Reuters EOD (≈ 24%), and Markit Intraday (≈ 11%) account for the majority of the data points, with numbers inside the brackets indicating their shares of the observations. Keep only one type of data contributor for each bank based on the number of observations. After this step, 8 different data contributors are left, with GFI FENICS accounting for around 78% of the data and Thomson Reuters EOD for around 14%. 289 banks are left and 205 are in EU or OECD countries.

Out of these 205 banks, 174 can be matched to Bankscope. Using ISIN number and Ticker can only match a limited number of banks since some banks are unlisted. So I manually match the banks from EIKON to the identifier (bvdid) in Bankscope using bank names, ISIN number and Ticker.

B.4.2 Data Sources

Table B.2: Data Sources

Data	Descriptions	Source
Concentration measures	HHI and 5-bank concentration ratio based on total assets of credit institutions	ECB
Concentration measures	HHI and 5-bank concentration ratio based on total assets of 6 types of banks	Bankscope annual statements, own calculation
Monetary and financial institutions (MFI) interest rates	harmonised monthly (annualised) lending rates and deposit rates on new business with an initial rate fixation period of 1 year	ECB
Credit default swap spreads	5-year CDS quarterly end spreads	Thomson Reuters EIKON
Quarterly bank-level variables	total assets, total equity	Bankscope quarterly statements
Annual bank-level variables	total assets, total equity, loan impairment charge, net income, etc.	Bankscope annual statements
Country-level macro variables	real GDP growth rate, inflation rate (growth rate of GDP deflator)	World Bank
Country-level macro variables	quarterly real GDP growth rate	OECD
Country-level total credit	total credit of domestic banks to private non-financial sector	BIS
Country-level total assets of credit institutions	total assets (in euros) of credit institutions including domestic banking groups and stand alone banks, foreign (EU and non-EU) controlled subsidiaries and foreign (EU and non-EU) controlled branches	ECB
Dollar/Euro exchange rate	used to convert the total assets of credit institutions in euros from the ECB into dollars	Federal Reserve Bank of St Louis (FRED)

B.4.3 Summary Statistics

Annual Bankscope Data

Table B.3: Summary Statistics of Key Variables by Groups of Countries

			Percentiles				
	Mean	Median	1st	25th	75th	99th	Obs.
EU countries							
change in equity/lagged assets	0.01	0.00	-0.08	-0.00	0.01	0.17	50,482
loan impairment ratio	0.01	0.01	-0.03	0.00	0.01	0.08	51,326
GDP growth rate	0.01	0.02	-0.06	0.00	0.03	0.07	56,942
inflation rate	0.02	0.01	-0.00	0.01	0.02	0.07	56,942
HHI (ECB)	0.11	0.10	0.02	0.06	0.12	0.40	56,307
HHI (Bankscope)	0.17	0.15	0.04	0.10	0.21	0.52	56,942
5-bank ratio (ECB)	0.59	0.59	0.20	0.47	0.71	0.99	56,350
5-bank ratio (Bankscope)	0.72	0.74	0.34	0.61	0.85	1.00	56,934
OECD countries							
change in equity/lagged assets	0.01	0.00	-0.06	0.00	0.01	0.15	209,680
loan impairment ratio	0.00	0.00	-0.01	0.00	0.01	0.06	221,648
GDP growth rate	0.02	0.02	-0.05	0.01	0.03	0.05	232,203
inflation rate	0.02	0.02	-0.01	0.01	0.02	0.05	232,203
HHI (Bankscope)	0.16	0.14	0.01	0.08	0.20	0.52	232,203
5-bank ratio (Bankscope)	0.69	0.72	0.20	0.57	0.83	1.00	232,195

Note: The table shows the summary statistics of the key variables used in the regression of change in equity on concentration. Change in total equity over lagged total assets is the dependent variable. Loan impairment ratio is calculated as loan impairment charge over gross loans. Variables apart from the concentration measures and GDP growth rate are winsorized for the top and bottom 1% of the distribution by a given group of countries (i.e., variables in the upper part of the table are winsorized by the pooled sample of EU countries). Statistics are shown after the winsorization. Statistics for concentration measures (country-year level) are computed using country-year level data, although bank-year observations are shown under "Obs.".

Table B.4: Data Description for EU/OECD Countries in Bankscope

Country	Period	Obs	Obs/Year	Commercial	Savings	Cooperative	BHC	Other
Australia	2005-2014	394	39	31	1	7	5	31
Austria	2000-2014	3,599	240	86	141	119	8	30
Belgium	1999-2014	963	60	53	16	9	11	24
Bulgaria	1999-2014	317	20	26	1	1	1	4
Canada	2010-2014	417	83	44	4	36	7	13
Chile	1999-2007	206	23	29	0	0	0	2
Croatia	1999-2014	567	35	50	1	1	0	16
Cyprus	1999-2014	225	14	22	1	2	4	0
Czech Republic	2004-2014	302	27	24	0	2	0	17
Denmark	1999-2014	1,654	103	66	60	10	5	20
Estonia	1999-2014	84	5	8	0	0	1	0
Finland	2005-2014	270	27	30	16	2	2	11
France	1999-2014	4,998	312	200	49	131	9	150
Germany	1999-2014	27,244	1,703	200	685	1,502	17	141
Greece	2005-2014	176	18	19	1	1	1	4
Hungary	1999-2014	411	26	38	1	1	0	22
Iceland	2003-2014	159	13	6	24	0	0	10
Ireland	2002-2014	253	19	21	0	0	4	27
Israel	1999-2014	187	12	17	0	0	1	3
Italy	2005-2014	6,282	628	143	53	508	14	66
Japan	1999-2014	10,621	664	175	0	673	33	65
Latvia	1999-2014	304	19	24	0	0	0	1
Lithuania	1999-2014	156	10	13	0	0	0	0
Luxembourg	1999-2014	1,229	77	132	2	2	11	11
Malta	2000-2014	131	9	12	1	1	0	2
Mexico	1999-2014	1,127	70	63	11	5	18	101
Netherlands	1999-2014	740	46	48	2	1	20	30
New Zealand	2006-2014	167	19	15	1	6	2	10
Norway	2006-2014	1,318	146	22	124	0	5	28
Poland	2004-2014	407	37	54	1	1	2	8
Portugal	2005-2014	648	65	28	84	4	7	14
Romania	1999-2014	312	20	31	3	1	0	7
Slovakia	2005-2014	142	14	14	2	0	1	6
Slovenia	2005-2014	183	18	15	2	2	0	5
South Korea	2010-2014	213	43	16	6	1	5	29
Spain	2005-2014	1,642	164	71	66	82	5	27
Sweden	1999-2014	1,387	87	34	81	1	10	25
Switzerland	1999-2014	5,372	336	216	251	10	28	24
Turkey	2006-2014	576	64	34	0	0	5	68
United Kingdom	2005-2014	2,316	232	145	4	1	44	143
United States	1999-2014	156,212	9,763	9,220	1,037	40	3,048	98

Note: "Obs" shows the total number of observations in a sample of six types of banks (i.e., bank holding companies, commercial banks, savings banks, cooperative banks, finance companies, and real estate & mortgage banks) for each country. "Obs/Year" shows the average number of observations in each year across the period covered in each country. The last five columns show the number of banks under each type category (i.e., commercial banks, savings banks, cooperative banks, bank holding companies and others). "Other" includes the other two types of banks.

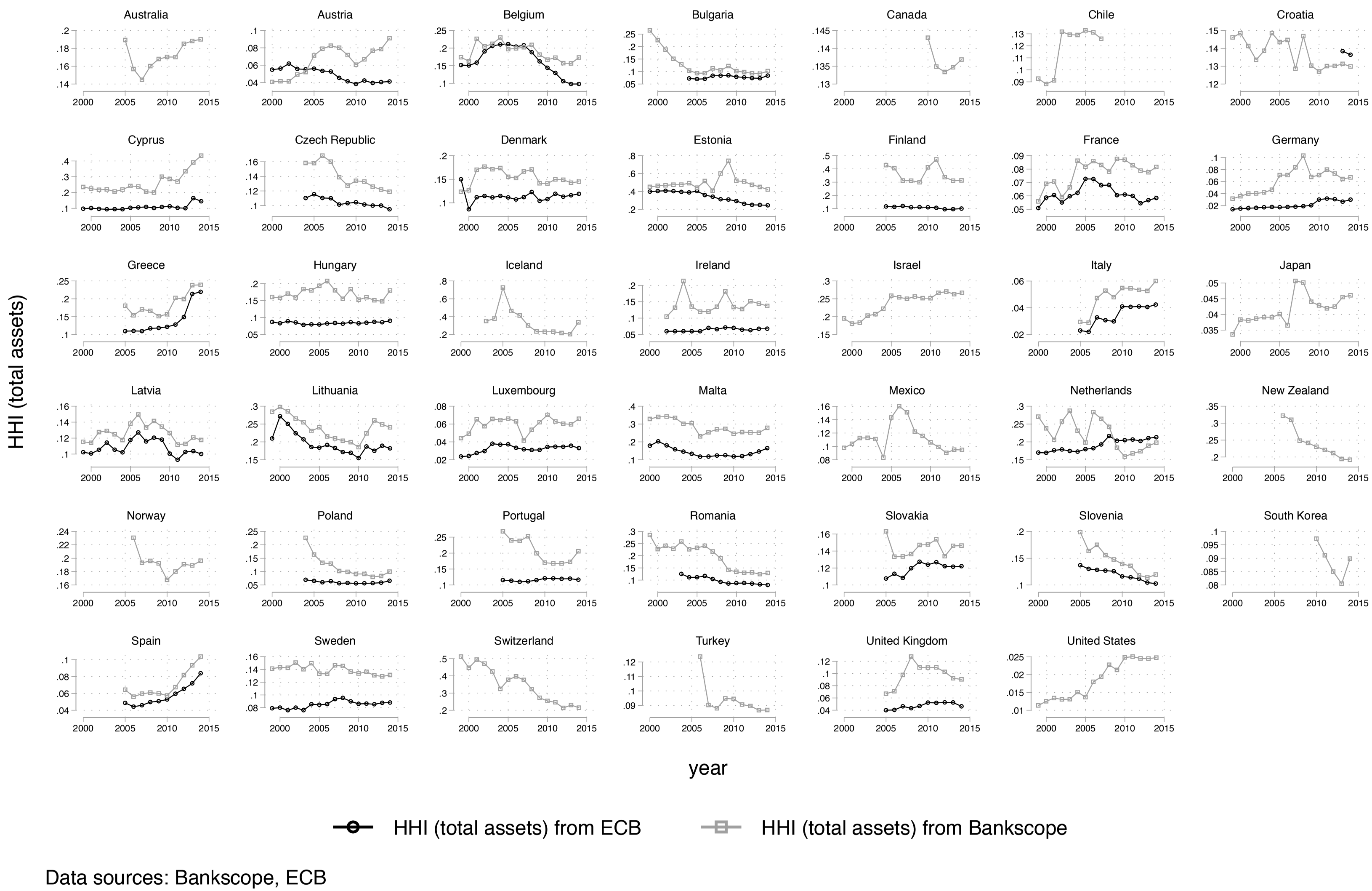
Table B.5: Bankscope Data Compared with the Aggregates from ECB/BIS

Country	Share of Total Assets (ECB)	Period	Share of Total Credit (BIS)	Period
Australia		-	0.96	2005-2014
Austria	0.62	2008-2014	1.13	2000-2014
Belgium	1.12	2007-2014	2.25	1999-2014
Bulgaria	0.85	2007-2014		-
Canada		-	1.29	2010-2014
Chile		-	1.04	1999-2007
Croatia	0.95	2013-2014		-
Cyprus	0.61	2008-2014		-
Czech Republic	0.90	2007-2014	1.13	2004-2014
Denmark	1.13	2008-2014	1.31	1999-2014
Estonia	0.48	2008-2014		-
Finland	1.08	2007-2014	1.17	2005-2014
France	1.49	2007-2014	1.60	1999-2014
Germany	0.93	2008-2014	1.33	1999-2014
Greece	0.75	2008-2014	0.92	2005-2014
Hungary	0.63	2008-2014	0.72	1999-2014
Iceland		-		-
Ireland	0.60	2008-2014	1.19	2002-2014
Israel		-	0.95	1999-2014
Italy	1.04	2007-2014	1.21	2005-2014
Japan		-	1.25	1999-2014
Latvia	0.85	2008-2014		-
Lithuania	0.84	2007-2014		-
Luxembourg	0.72	2008-2014	4.76	2003-2014
Malta	0.36	2007-2014		-
Mexico		-	1.76	1999-2014
Netherlands	0.84	2008-2014	0.86	1999-2014
New Zealand		-	0.88	2006-2014
Norway		-	1.10	2006-2014
Poland	0.72	2007-2014	0.81	2004-2014
Portugal	0.94	2007-2014	0.96	2005-2014
Romania	0.75	2007-2014		-
Slovakia	0.71	2007-2014		-
Slovenia	0.82	2007-2014		-
South Korea		-	0.65	2010-2014
Spain	0.77	2008-2014	1.04	2005-2014
Sweden	0.64	2007-2014	0.99	1999-2014
Switzerland		-	1.09	1999-2014
Turkey		-	1.32	2006-2014
United Kingdom	0.75	2008-2014	1.49	2005-2014
United States		-	2.01	1999-2014

Data sources: Bankscope, ECB, BIS, FRED

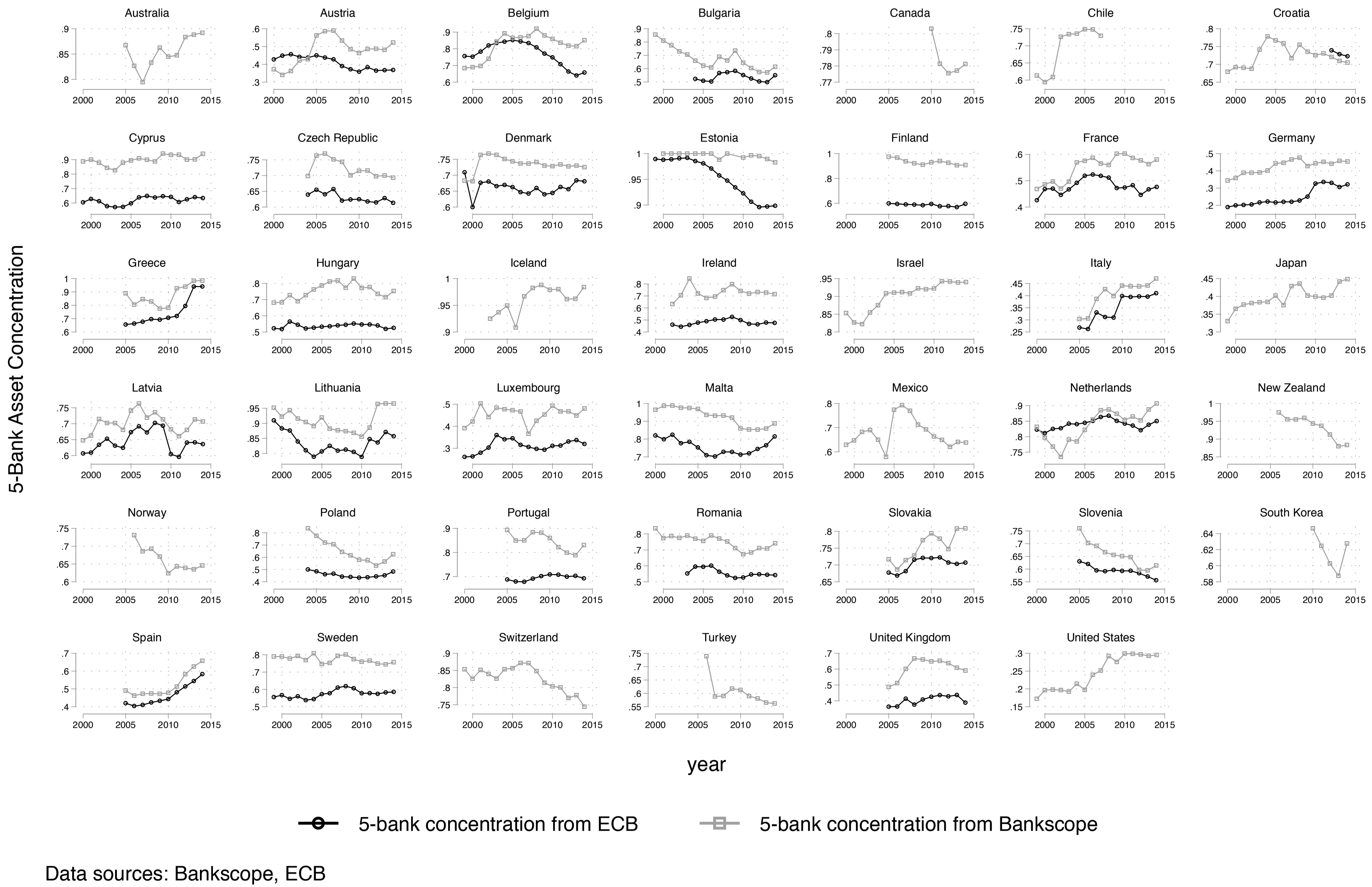
Note: Share of total assets (ECB) is computed by dividing total assets of all sampled banks in Bankscope data by total assets of credit institutions from ECB. The numbers reported are mean values over the period indicated in the third column. Share of total credit (BIS) is computed by dividing total gross loans of all sampled banks in Bankscope by total credit of domestic banks to private non-financial sector from BIS. The numbers reported are mean values over the period indicated in the last column.

Figure B.1: Concentration Index (HHI) for EU and OECD Countries from 1999 to 2014



Note: The graphs plot the Herfindahl Hirschman Index (HHI) over time for each EU/OECD country. For EU countries where the measure is available from the ECB, the black line corresponds to the ECB measure over time. HHI from ECB is based on total assets of credit institutions, while HHI from Bankscope (grey line) is computed using total assets of 6 types of banks (i.e., bank holding companies, commercial banks, cooperative banks, savings banks, finance companies and real estate & mortgage banks).

Figure B.2: 5-Bank Asset Concentration Ratio for EU and OECD Countries from 1999 to 2014



Note: The graphs plot the 5-bank asset concentration ratio over time for each EU/OECD country. For EU countries where the measure is available from the ECB, the black line corresponds to the ECB measure over time. 5-bank concentration ratio from ECB is based on total assets of credit institutions, while the concentration ratio from Bankscope (grey line) is computed using total assets of 6 types of banks (i.e., bank holding companies, commercial banks, cooperative banks, savings banks, finance companies and real estate & mortgage banks).

Merged Sample of Quarterly CDS Spreads and Bankscope Data

Table B.6: Description for the Merged Sample
(quarterly CDS data merged with quarterly Bankscope data)

Country	Period	Obs	Obs/Year	Banks	Q1	Q2	Q3	Q4
Australia	2005-2016	154	13	8	0	68	3	83
Austria	2004-2016	126	10	5	25	36	29	36
Belgium	2005-2016	43	4	3	0	20	0	23
Canada	2010-2016	151	22	6	36	36	37	42
Chile	2011-2016	23	4	1	6	6	6	5
Denmark	2004-2016	49	4	1	13	12	12	12
Finland	2008-2016	17	2	1	0	9	0	8
France	2004-2016	230	18	11	22	91	25	92
Germany	2003-2016	208	15	7	44	62	43	59
Greece	2008-2016	108	12	4	26	28	27	27
Hungary	2008-2016	29	3	1	7	7	7	8
Ireland	2004-2016	86	7	4	0	42	1	43
Italy	2005-2016	234	20	6	48	62	58	66
Japan	2003-2016	368	26	19	53	134	49	132
Netherlands	2004-2016	133	10	7	11	56	4	62
Norway	2006-2016	81	7	2	19	22	20	20
Portugal	2004-2016	126	10	4	28	37	26	35
South Korea	2009-2016	157	20	7	34	41	33	49
Spain	2004-2016	260	20	9	53	72	57	78
Sweden	2005-2016	202	17	5	50	52	51	49
Switzerland	2003-2016	99	7	2	25	25	24	25
Turkey	2007-2016	86	9	4	15	26	20	25
United Kingdom	2004-2016	264	20	14	21	102	22	119
United States	2003-2016	983	70	26	239	244	248	252

Note: The table shows the number of observations for the merged sample of banks in OECD/EU countries. "Period" shows the time coverage for each country. "Obs" shows the total number of observations for each country. "Banks" shows the number of banks present in the sample. Columns Q1-Q4 show the number of total observations in each quarter.

Table B.7: Summary Statistics of Key Variables in the Merged Sample by Country

	CDS (decimals)			Equity Ratio			
Country	1st	50th	99th	1st	50th	99th	Obs.
Australia	0.00	0.01	0.02	0.02	0.06	0.15	154
Austria	0.00	0.01	0.04	0.03	0.07	0.09	126
Belgium	0.00	0.01	0.05	0.03	0.05	0.09	43
Canada	0.00	0.01	0.04	0.03	0.05	0.06	151
Chile	0.01	0.01	0.02	0.07	0.09	0.09	23
Denmark	0.00	0.01	0.03	0.03	0.03	0.05	49
Finland	0.00	0.01	0.01	0.03	0.04	0.06	17
France	0.00	0.01	0.05	-0.00	0.03	0.10	230
Germany	0.00	0.01	0.04	0.01	0.03	0.07	208
Greece	0.01	0.09	0.26	-0.03	0.06	0.13	108
Hungary	0.01	0.01	0.02	0.09	0.12	0.15	29
Ireland	0.00	0.04	0.23	-0.01	0.04	0.12	86
Italy	0.00	0.01	0.06	0.03	0.07	0.14	234
Japan	0.00	0.01	0.04	0.02	0.06	0.09	368
Netherlands	0.00	0.01	0.18	0.02	0.04	0.34	133
Norway	0.00	0.01	0.02	0.04	0.05	0.07	81
Portugal	0.00	0.04	0.18	0.03	0.05	0.10	126
South Korea	0.01	0.01	0.02	0.05	0.07	0.12	157
Spain	0.00	0.02	0.10	0.02	0.06	0.08	260
Sweden	0.00	0.01	0.03	0.02	0.04	0.06	202
Switzerland	0.00	0.01	0.03	0.01	0.04	0.06	99
Turkey	0.02	0.03	0.08	0.10	0.12	0.15	86
United Kingdom	0.00	0.01	0.05	0.01	0.05	0.08	264
United States	0.00	0.01	0.08	0.05	0.10	0.20	983

Note: The table shows the summary statistics CDS spreads and equity-to-total assets ratio for each country in the merged sample. Numbers reported are in decimal places, e.g., CDS spread of 0.01 refers to 100 basis points, equity ratio of 0.02 means 20%. For each variable, the 1st, 50th, 99th percentiles are reported. "Obs." shows the number of bank-quarter observations in each country.

B.5 Robustness Checks

Table B.8: The Effect of Concentration (5-Bank Asset Concentration Ratio) on Change in Total Equity over Lagged Total Assets during 1999-2014

	(1) EU	(2) EU	(3) EU	(4) EU	(5) OECD	(6) OECD
L.5-bank ratio (ECB)	0.06***	0.04***				
	(0.00)	(0.00)				
L.5-bank ratio (Bankscope)			-0.01	-0.00	0.02***	0.03***
			(0.00)	(0.00)	(0.00)	(0.00)
L.loan impairment ratio		-0.06***		-0.06***		-0.16***
		(0.02)		(0.02)		(0.01)
L.GDP growth rate		0.10***		0.13***		0.06***
		(0.01)		(0.01)		(0.01)
inflation rate		0.09***		0.12***		0.15***
		(0.02)		(0.02)		(0.01)
Observations	44,499	44,499	45,026	45,026	199,310	199,310
No.banks	4,915	4,915	4,936	4,936	19,230	19,230
Adjusted R^2	0.275	0.281	0.264	0.274	0.104	0.111
Within R^2	0.010	0.018	0.000	0.014	0.000	0.008
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Bank-level clustered standard errors in parentheses

Data sources: Bankscope annual data, ECB, World Bank

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

Note: The table shows the results from regressing change in total equity over lagged total assets on lagged 5-bank asset concentration ratio and lagged loan impairment ratio (computed as loan impairment charge/gross loans), controlling for lagged real GDP growth and inflation rate (i.e., growth rate of GDP deflator). 5-bank ratio (ECB) is the ECB estimate of 5-bank asset concentration based on the total assets of credit institutions in EU countries. 5-bank ratio (Bankscope) is calculated using 6 types of banks (i.e., bank holding companies, commercial banks, cooperative banks, finance companies, real estate & mortgage banks, and savings banks) from annual Bankscope data.

Table B.9: The Effect of Concentration (Herfindahl Hirschman Index HHI) on Change in Total Equity over Lagged Total Assets during 1999-2014

	(1) Euro	(2) Euro	(3) non-Euro	(4) non-Euro	(5) Euro	(6) Euro	(7) non-Euro	(8) non-Euro	(9) non-EU	(10) non-EU
L.HHI (ECB)	0.13***	0.11***	0.49***	0.29***						
	(0.02)	(0.02)	(0.13)	(0.11)						
L.HHI (Bankscope)					-0.02*	-0.03**	0.12***	0.15***	0.03***	0.04***
					(0.01)	(0.01)	(0.04)	(0.04)	(0.00)	(0.00)
L.loan impairment ratio		-0.06***		0.01		-0.06***		-0.01		-0.21***
		(0.02)		(0.07)		(0.02)		(0.05)		(0.01)
L.GDP growth rate		0.08***		0.04		0.09***		0.13***		0.07***
		(0.01)		(0.04)		(0.01)		(0.04)		(0.01)
inflation rate		0.16***		0.27***		0.17***		-0.02		-0.01
		(0.02)		(0.06)		(0.02)		(0.05)		(0.02)
Observations	40,060	40,060	4,359	4,359	40,060	40,060	4,973	4,973	155,634	155,634
No.banks	4,307	4,307	568	568	4,307	4,307	629	629	14,449	14,449
Adjusted R^2	0.277	0.286	0.325	0.333	0.274	0.284	0.296	0.298	0.098	0.103
Within R^2	0.004	0.017	0.008	0.021	0.000	0.015	0.005	0.008	0.001	0.007
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Bank-level clustered standard errors in parentheses

Data sources: Bankscope annual data, ECB, World Bank

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

Note: The table shows the results from regressing change in total equity over lagged total assets on lagged concentration index HHI and lagged loan impairment ratio (computed as loan impairment charge/gross loans), controlling for lagged real GDP growth and inflation rate (i.e., growth rate of GDP deflator). HHI (ECB) is the ECB estimate of HHI based on the total assets of credit institutions in EU countries. HHI (Bankscope) is calculated using 6 types of banks (i.e., bank holding companies, commercial banks, cooperative banks, finance companies, real estate & mortgage banks, and savings banks) from annual Bankscope data. The samples are split into different groups of countries. Euro refers to Eurozone countries. Non-Euro refers to non-eurozone EU countries and non-EU refers to non-EU OECD countries.

Table B.10: The Effect of Concentration (Herfindahl Hirschman Index HHI) on Pre-dividend Change in Equity over Lagged Total Assets during 1999-2014

	(1) EU	(2) EU	(3) EU	(4) EU	(5) OECD	(6) OECD
L.HHI (ECB)	0.16***	0.12***				
	(0.02)	(0.02)				
L.HHI (Bankscope)			0.06***	0.04***	0.05***	0.04***
			(0.01)	(0.01)	(0.00)	(0.00)
L.loan impairment ratio		-0.07***		-0.07***		-0.19***
		(0.02)		(0.02)		(0.01)
L.GDP growth rate		0.11***		0.11***		0.05***
		(0.01)		(0.01)		(0.01)
inflation rate		0.13***		0.13***		0.14***
		(0.02)		(0.02)		(0.01)
Observations	44,340	44,340	44,950	44,950	199,129	199,129
No.banks	4,870	4,870	4,930	4,930	19,223	19,223
Adjusted R^2	0.290	0.298	0.285	0.295	0.178	0.184
Within R^2	0.004	0.016	0.002	0.015	0.001	0.009
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Bank-level clustered standard errors in parentheses

Data sources: Bankscope annual data, ECB, World Bank

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

Note: The table shows the results from regressing pre-dividend change in equity (i.e., change in equity plus the cash dividends) over lagged total assets on lagged concentration index HHI and lagged loan impairment cost (computed as loan impairment charge/gross loans), controlling for lagged real GDP growth and inflation rate (i.e., growth rate of GDP deflator). HHI (ECB) is the ECB estimate of HHI based on the total assets of credit institutions in EU countries. HHI (Bankscope) is calculated using 6 types of banks (i.e., bank holding companies, commercial banks, cooperative banks, finance companies, real estate & mortgage banks, and savings banks) from annual Bankscope data.

Table B.11: The Effect of Concentration (Herfindahl Hirschman Index HHI) on Change in Equity over Lagged Total Assets during 1999-2014

	(1) EU 1999-2006	(2) EU 2006-2014	(3) EU 2010-2014	(4) EU 1999-2006	(5) EU 2006-2014	(6) EU 2010-2014
L.HHI (ECB)	0.01	0.18***	0.13***	0.01	0.14***	0.12***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.02)	(0.04)
L.loan impairment ratio				-0.03	-0.05**	0.13***
				(0.03)	(0.02)	(0.02)
L.GDP growth rate				0.02	0.09***	0.01
				(0.04)	(0.01)	(0.02)
inflation rate				-0.01	0.16***	-0.04
				(0.04)	(0.02)	(0.04)
Observations	16,771	30,970	17,176	16,771	30,970	17,176
No.banks	3,111	4,322	3,818	3,111	4,322	3,818
Adjusted R^2	0.350	0.243	0.220	0.350	0.253	0.226
Within R^2	0.000	0.004	0.001	0.000	0.018	0.009
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Bank-level clustered standard errors in parentheses

Data sources: Bankscope annual data, ECB, World Bank

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

Note: The table shows the results from regressing change in total equity over lagged total assets on lagged concentration index HHI and lagged loan impairment cost (computed as loan impairment charge/gross loans), controlling for lagged real GDP growth and inflation rate (i.e., growth rate of GDP deflator). HHI (ECB) is the ECB estimate of HHI based on the total assets of credit institutions in EU countries. The sample is divided into different subgroups conditioning on time period.

Table B.12: The Effect of Bank Equity Ratio on Bank CDS Spread during Different Time Periods

	(1) EU 2003-2011	(2) EU 2011-2016	(3) Eurozone 2003-2011	(4) Eurozone 2011-2016	(5) OECD 2003-2011	(6) OECD 2011-2016
L.equity ratio	-0.14	-0.34***	-0.13	-0.32**	-0.38**	-0.33***
	(0.16)	(0.12)	(0.17)	(0.12)	(0.17)	(0.10)
L.loan impairment ratio	2.40***	0.26	2.94**	0.34*	0.99**	0.18
	(0.83)	(0.17)	(1.12)	(0.18)	(0.39)	(0.16)
L.GDP growth rate	-0.64***	-0.95***	-0.84***	-1.35***	-0.44**	-0.52***
	(0.19)	(0.25)	(0.17)	(0.26)	(0.18)	(0.15)
Observations	582	862	434	636	1,195	1,933
Number of Banks	38	47	29	35	83	101
Adjusted R^2	0.685	0.824	0.708	0.831	0.641	0.819
Within R^2	0.164	0.197	0.181	0.236	0.153	0.163
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Bank-level clustered standard errors in parentheses

Data sources: Thomson Reuters EIKON, Bankscope quarterly data, OECD

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

Note: The table shows the results from regressing 5-year CDS spreads on banks' equity ratios, controlling for loan impairment charge to gross loans ratios, and real GDP growth rate. Bank, country and quarter fixed effects are included in all regressions. Quarterly data are used and all variables are in decimal places. Lagged explanatory variables are used. The sample is divided into different subgroups according to regions and time periods.

Table B.13: The Effect of Bank Equity Ratio on Bank CDS Spread during 2003-2016 Using Annual Data

	(1) EU	(2) EU	(3) Eurozone	(4) Eurozone	(5) OECD	(6) OECD
L.equity ratio	-0.21***	-0.05	-0.20***	-0.02	-0.20***	-0.08*
	(0.07)	(0.06)	(0.07)	(0.07)	(0.05)	(0.04)
L.loan impairment ratio		0.43*		0.49*		0.48**
		(0.25)		(0.28)		(0.22)
L.GDP growth rate		-0.29**		-0.34**		-0.16**
		(0.11)		(0.14)		(0.07)
Observations	628	627	461	461	1,181	1,145
Number of Banks	74	74	54	54	144	141
Adjusted R^2	0.634	0.675	0.640	0.684	0.632	0.667
Within R^2	0.028	0.140	0.025	0.149	0.041	0.135
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Bank-level clustered standard errors in parentheses

Data sources: Thomson Reuters EIKON, Bankscope annual data, World Bank

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

Note: The table shows the results from regressing 5-year CDS spreads on banks' equity ratios, controlling for loan impairment charge to gross loans ratios, and real GDP growth rate. Bank, country and year fixed effects are included in all regressions. Annual data are used and all variables are in decimal places. Lagged explanatory variables are used. The sample is divided into different subgroups according to regions.

Table B.14: The Effect of Bank Equity Ratio on Bank CDS Spread during 2003-2016 Using Different Fixed Effects

	(1) EU	(2) EU	(3) Eurozone	(4) Eurozone	(5) OECD	(6) OECD
L.equity ratio	-0.14*	-0.11*	-0.15*	-0.11*	-0.20**	-0.25*
	(0.07)	(0.06)	(0.07)	(0.06)	(0.08)	(0.13)
L.loan impairment ratio		0.13		0.14		0.16**
		(0.11)		(0.12)		(0.08)
L.GDP growth rate		-0.24***		-0.28***		-0.14***
		(0.05)		(0.06)		(0.05)
Observations	1,343	1,339	997	993	3,001	2,864
Number of Banks	50	50	38	38	107	103
Adjusted R^2	0.861	0.863	0.856	0.859	0.821	0.828
Within R^2	0.012	0.031	0.012	0.032	0.046	0.059
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country*Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Bank-level clustered standard errors in parentheses

Data sources: Thomson Reuters EIKON, Bankscope quarterly data, OECD

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

Note: The table shows the results from regressing 5-year CDS spreads on banks' equity ratios, controlling for loan impairment charge to gross loans ratios, and real GDP growth rate. Bank fixed effects and country*year fixed effects are included in all regressions. Quarterly data are used and all variables are in decimal places. Lagged explanatory variables are used. The sample is divided into different subgroups according to regions.

Table B.15: Direct Relationship between CDS Spread and Concentration Measures in EU Countries

	(1) EU 2003-2016	(2) EU 2003-2011	(3) EU 2011-2016	(4) EU 2003-2016	(5) EU 2003-2011	(6) EU 2011-2016
L.HHI (Bankscope)	-0.06	-0.03	-0.30**			
	(0.07)	(0.10)	(0.12)			
L.equity ratio	-0.06	-0.36*	0.03	-0.07	-0.37*	-0.01
	(0.06)	(0.19)	(0.08)	(0.06)	(0.19)	(0.08)
L.loan impairment ratio	0.44*	1.12***	0.15	0.47*	1.11***	0.14
	(0.26)	(0.35)	(0.13)	(0.26)	(0.35)	(0.14)
L.GDP growth rate	-0.29**	-0.30**	-0.27***	-0.30***	-0.30**	-0.31***
	(0.11)	(0.15)	(0.09)	(0.11)	(0.14)	(0.09)
L.5-bank ratio (Bankscope)				-0.06***	-0.04**	-0.11**
				(0.02)	(0.02)	(0.04)
Observations	621	336	350	621	336	350
Number of Banks	74	65	74	74	65	74
Adjusted R^2	0.675	0.605	0.849	0.682	0.608	0.845
Within R^2	0.145	0.252	0.210	0.162	0.257	0.185
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Bank-level clustered standard errors in parentheses

Data sources: Thomson Reuters EIKON, Bankscope annual data, World Bank

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

Note: The table shows the results from regressing 5-year CDS spreads on concentration index HHI or 5-bank concentration ratio, controlling for banks' equity ratios, loan impairment charge to gross loans ratios, and real GDP growth rate and including bank, country and year fixed effects. Annual data are used and all variables are in decimals. Lagged explanatory variables are used. The sample consists of EU banks and is divided into different sub-samples based on time periods.

Appendix C

Appendix to Chapter 4

C.1 Solving Firm's Problem

Let $\lambda_{i,t}$ denote the Lagrange multiplier associated with the borrowing constraint (4.10) of firm i in period t. Using (4.7), (4.8), (4.9) and (4.10), form the Lagrangian:

$$\begin{aligned}\mathcal{L} =& \mathrm{E}_t \sum_{\tau=0}^{\infty} \varphi(1-\varphi)^{\tau}\eta^{\tau} \{p_{i,t+1+\tau}y_{i,t+1+\tau} - w_{t+1+\tau}l_{i,t+1+\tau} - p_{m,t+1+\tau}m_{i,t+1+\tau} \\ & -R_{t+\tau}(k_{i,t+\tau} - n_{i,t+\tau}) + (1-\delta)k_{i,t+\tau} + \lambda_{i,t+\tau}\left[\frac{n_{i,t+\tau}}{1-\phi(1-\delta)} - k_{i,t+\tau}\right]\right\}\end{aligned} \tag{C.1}$$

Using (4.4) and (C.1), the first-order conditions with respect to the firm's capital demand $k_{i,t}$, labor demand $l_{i,t}$ and materials demand $m_{i,t}$ are respectively:

$$\mathrm{E}_t\left[\beta_k \frac{p_{i,t+1}y_{i,t+1}}{k_{i,t}} - R_t + (1-\delta)\right] = \lambda_{i,t} \tag{C.2}$$

$$\beta_l \frac{p_{i,t}y_{i,t}}{l_{i,t}} = w_t \tag{C.3}$$

$$\beta_m \frac{p_{i,t}y_{i,t}}{m_{i,t}} = p_{m,t} \tag{C.4}$$

The nominal value added $\mathrm{VA}_{i,t}$ is equivalent to a fraction of the nominal revenue $p_{i,t}y_{i,t}$, using (4.4) and (C.4):

$$\mathrm{VA}_{i,t} \equiv p_{i,t}y_{i,t} - p_{m,t}m_{i,t} = (1-\beta_m)p_{i,t}y_{i,t} \tag{C.5}$$

C.1.1 Marginal Revenue Product of Capital

Divide (C.3) by (C.4) to get the materials-to-labor ratio:

$$\frac{m_{i,t}}{l_{i,t}} = \frac{\beta_m w_t}{\beta_l p_{m,t}} \tag{C.6}$$

Use (4.4) and the first order condition with respect to labor (C.3) to write the optimal labor demand $l_{i,t+1}$ in terms of w_{t+1}, $Z_{i,t+1}$, $k_{i,t}$ and $m_{i,t+1}$:

$$l_{i,t+1} = \left[\frac{w_{t+1}}{\beta_l Z_{i,t+1} k_{i,t}^{\beta_k} m_{i,t+1}^{\beta_m}}\right]^{\frac{1}{\beta_l - 1}} = \left(\frac{\beta_l}{w_{t+1}}\right)^{\frac{1}{1-\beta_l}} Z_{i,t+1}^{\frac{1}{1-\beta_l}} k_{i,t}^{\frac{\beta_k}{1-\beta_l}} m_{i,t+1}^{\frac{\beta_m}{1-\beta_l}} \tag{C.7}$$

Use (C.6) and (C.7) to get:

$$l_{i,t+1} = \left(\frac{\beta_l}{w_{t+1}}\right)^{\frac{1}{1-\beta_l}} Z_{i,t+1}^{\frac{1}{1-\beta_l}} k_{i,t}^{\frac{\beta_k}{1-\beta_l}} \left(\frac{\beta_m w_{t+1}}{\beta_l p_{m,t+1}} l_{i,t+1}\right)^{\frac{\beta_m}{1-\beta_l}} \tag{C.8}$$

Rearrange to get the optimal labor demand in terms of w_{t+1}, $p_{m,t+1}$, $Z_{i,t+1}$, and $k_{i,t}$:

$$l_{i,t+1} = \left(\frac{\beta_l}{w_{t+1}}\right)^{\frac{1-\beta_m}{1-\beta_l-\beta_m}} \left(\frac{\beta_m}{p_{m,t+1}}\right)^{\frac{\beta_m}{1-\beta_l-\beta_m}} Z_{i,t+1}^{\frac{1}{1-\beta_l-\beta_m}} k_{i,t}^{\frac{\beta_k}{1-\beta_l-\beta_m}} \tag{C.9}$$

Use (C.6) and (C.9) to write the optimal materials demand in terms of w_{t+1}, $p_{m,t+1}$, $Z_{i,t+1}$, and $k_{i,t}$:

$$m_{i,t+1} = \left(\frac{\beta_l}{w_{t+1}}\right)^{\frac{\beta_l}{1-\beta_l-\beta_m}} \left(\frac{\beta_m}{p_{m,t+1}}\right)^{\frac{1-\beta_l}{1-\beta_l-\beta_m}} Z_{i,t+1}^{\frac{1}{1-\beta_l-\beta_m}} k_{i,t}^{\frac{\beta_k}{1-\beta_l-\beta_m}} \tag{C.10}$$

The expressions for labor and materials demand hold for both unconstrained and constrained firms. Substitute (C.9) and (C.10) into the production function (4.4) to write $p_{i,t+1}y_{i,t+1}$ in terms of w_{t+1}, $p_{m,t+1}$, $Z_{i,t+1}$, and $k_{i,t}$:

$$p_{i,t+1}y_{i,t+1} = \left(\frac{\beta_l}{w_{t+1}}\right)^{\frac{\beta_l}{1-\beta_l-\beta_m}} \left(\frac{\beta_m}{p_{m,t+1}}\right)^{\frac{\beta_m}{1-\beta_l-\beta_m}} Z_{i,t+1}^{\frac{1}{1-\beta_l-\beta_m}} k_{i,t}^{\frac{\beta_k}{1-\beta_l-\beta_m}} \tag{C.11}$$

Use $\text{MRPK}_{i,t} = \beta_k \frac{p_{i,t}y_{i,t}}{k_{i,t-1}}$ (4.14), (C.11), and $\beta_k + \beta_l + \beta_m = \frac{\epsilon-1}{\epsilon}$ to write the log of the marginal revenue product of capital in terms of w_t, $p_{m,t}$, $Z_{i,t}$, and $k_{i,t-1}$:

$$\begin{aligned} \ln\text{MRPK}_{i,t} =& \frac{\epsilon}{1+\epsilon\beta_k}\ln Z_{i,t} - \frac{1}{1+\epsilon\beta_k}\ln k_{i,t-1} - \frac{\epsilon\beta_l}{1+\epsilon\beta_k}\ln w_t - \frac{\epsilon\beta_m}{1+\epsilon\beta_k}\ln p_{m,t} \\ &+ \ln\left(\beta_k \beta_l^{\frac{\epsilon\beta_l}{1+\epsilon\beta_k}} \beta_m^{\frac{\epsilon\beta_m}{1+\epsilon\beta_k}}\right) \end{aligned} \tag{C.12}$$

C.1.2 Dispersion of MRPK across All Firms

Using (C.12), the cross-section dispersion of the marginal revenue product of capital $\text{Var}_i(\text{lnMRPK}_{i,t})$ within a given industry in period t can be written as:

$$\begin{aligned}\text{Var}_i(\text{lnMRPK}_{i,t}) =& \left(\frac{\epsilon}{1+\epsilon\beta_k}\right)^2 \text{Var}_i(\text{ln}Z_{i,t}) + \left(\frac{1}{1+\epsilon\beta_k}\right)^2 \text{Var}_i(\text{ln}k_{i,t-1}) \\ & - 2\frac{\epsilon}{(1+\epsilon\beta_k)^2}\text{Cov}_i(\text{ln}Z_{i,t}, \text{ln}k_{i,t-1}) \\ =& \psi_1\text{Var}_i(\text{ln}Z_{i,t}) + \psi_2\text{Var}_i(\text{ln}k_{i,t-1}) - \psi_3\text{Cov}_i(\text{ln}Z_{i,t}, \text{ln}k_{i,t-1})\end{aligned} \tag{C.13}$$

where $\psi_1 \equiv \left(\frac{\epsilon}{1+\epsilon\beta_k}\right)^2$, $\psi_2 \equiv \left(\frac{\epsilon}{1+\epsilon\beta_k}\right)^2$, and $\psi_3 \equiv 2\frac{\epsilon}{(1+\epsilon\beta_k)^2}$. Rewrite the dispersion of MRPK in terms of the exogenous or predetermined variables by using $Z_{i,t} \equiv Z_t z_i z_{i,t}$ and the AR(1) process for $\text{ln}z_{i,t}$ (4.5):

$$\begin{aligned}\text{Var}_i(\text{lnMRPK}_{i,t}) =& \psi_1\text{Var}_i(\text{ln}Z_t + \text{ln}z_i + \rho\text{ln}z_{i,t-1} + e_{i,t}) + \psi_2\text{Var}_i(\text{ln}k_{i,t-1}) \\ & - \psi_3\text{Cov}_i(\text{ln}Z_t + \text{ln}z_i + \rho\text{ln}z_{i,t-1} + e_{i,t}, \text{ln}k_{i,t-1}) \\ =& \psi_1\text{Var}_i(\text{ln}z_i) + \psi_1\text{Var}_i(e_{i,t}) + \psi_1\rho^2\text{Var}_i(\text{ln}z_{i,t-1}) + \psi_2\text{Var}_i(\text{ln}k_{i,t-1}) \\ & - \psi_3\text{Cov}_i(\text{ln}z_i + \rho\text{ln}z_{i,t-1}, \text{ln}k_{i,t-1})\end{aligned} \tag{4.15}$$

This is a general decomposition and holds even in the presence of constrained firms. Since $k_{i,t}$ is driven by different processes for unconstrained and constrained firms, by replacing $\text{ln}k_{i,t}$ by $\text{ln}k^U_{i,t}$ (4.13) or $\text{ln}k^C_{i,t}$ (4.12), it is possible to find the dispersion of MRPK within the two subgroups of firms, i.e., $\text{Var}_i(\text{lnMRPK}^U_{i,t})$ and $\text{Var}_i(\text{lnMRPK}^C_{i,t})$.

C.1.3 Capital Demand of Financially Unconstrained Firms

If firm i is unconstrained in period t (i.e., $\lambda_{i,t} = 0$), then the first order condition (C.2) can be simplified to:

$$\text{E}_t\left[\beta_k\frac{p_{i,t+1}y_{i,t+1}}{k_{i,t}}\right] = R_t - (1-\delta) = r_t + \delta \tag{C.14}$$

where $r_t \equiv R_t - 1$ is the net real interest rate. Rearrange (C.14) to get the unconstrained capital demand in terms of the expected revenue:

$$k^U_{i,t} = \frac{\beta_k}{r_t + \delta}\text{E}_t\left[p_{i,t+1}y_{i,t+1}\right] \tag{C.15}$$

Hence,

$$\Delta\text{ln}k^U_{i,t} \equiv \text{ln}k^U_{i,t} - \text{ln}k^U_{i,t-1} = \text{ln}\frac{r_{t-1}+\delta}{r_t+\delta} + \Delta\text{lnE}_t\left[p_{i,t+1}y_{i,t+1}\right] \tag{C.16}$$

As can be seen above, the investment of an unconstrained firm is driven by the change in the net real interest rate and the growth in expected sales $p_{i,t+1}y_{i,t+1}$ or value added

$(1-\beta_m)p_{i,t+1}y_{i,t+1}$. Alternatively, substitute (C.11) into (C.15) to get:

$$k_{i,t}^U = \frac{\beta_k}{r_t+\delta}\mathrm{E}_t\left[\left(\frac{\beta_l}{w_{t+1}}\right)^{\frac{\beta_l}{1-\beta_l-\beta_m}}\left(\frac{\beta_m}{p_{m,t+1}}\right)^{\frac{\beta_m}{1-\beta_l-\beta_m}} Z_{i,t+1}^{\frac{1}{1-\beta_l-\beta_m}} k_{i,t}^{\frac{\beta_k}{1-\beta_l-\beta_m}}\right] \tag{C.17}$$

Rearrange to solve for the following optimal unconstrained capital demand chosen in period t:

$$(k_{i,t}^U)^{\frac{1-\beta_l-\beta_m-\beta_k}{1-\beta_l-\beta_m}} = \frac{\beta_k}{r_t+\delta}\mathrm{E}_t\left[\left(\frac{\beta_l}{w_{t+1}}\right)^{\frac{\beta_l}{1-\beta_l-\beta_m}}\left(\frac{\beta_m}{p_{m,t+1}}\right)^{\frac{\beta_m}{1-\beta_l-\beta_m}} Z_{i,t+1}^{\frac{1}{1-\beta_l-\beta_m}}\right] \tag{C.18}$$

Use $Z_{i,t+1} \equiv Z_{t+1}z_i z_{i,t+1}$ and the assumption that the idiosyncratic transitory productivity $z_{i,t+1}$ is independent from the trend Z_{t+1} or the idiosyncratic permanent productivity z_i, to get:

$$\begin{aligned}\ln k_{i,t}^U =&(1+\epsilon\beta_k)\left\{\ln\left(\beta_k\beta_l^{\frac{\epsilon\beta_l}{1+\epsilon\beta_k}}\beta_m^{\frac{\epsilon\beta_m}{1+\epsilon\beta_k}}\right)-\ln(r_t+\delta)\right.\\ &\left.+\ln\mathrm{E}_t\left[\left(\frac{Z_{t+1}}{w_{t+1}^{\beta_l}p_{m,t+1}^{\beta_m}}\right)^{\frac{\epsilon}{1+\epsilon\beta_k}}\right]+\ln\mathrm{E}_t\left[z_{i,t+1}^{\frac{\epsilon}{1+\epsilon\beta_k}}\right]+\frac{\epsilon}{1+\epsilon\beta_k}\ln z_i\right\}\end{aligned} \tag{C.19}$$

According to (4.5), the firm's productivity $z_{i,t+1}$ follows the AR(1) process:

$$z_{i,t+1} = z_{i,t}^{\rho}\exp(e_{i,t+1}) \tag{C.20}$$

where $e_{i,t+1} \overset{i.i.d.}{\sim} N(0,\sigma_z^2)$, so

$$\mathrm{E}_t\left[z_{i,t+1}^{\frac{\epsilon}{1+\epsilon\beta_k}}\right] = \mathrm{E}_t\left[z_{i,t}^{\frac{\rho\epsilon}{1+\epsilon\beta_k}}\exp\left(\frac{e_{i,t+1}\epsilon}{1+\epsilon\beta_k}\right)\right] = z_{i,t}^{\frac{\rho\epsilon}{1+\epsilon\beta_k}}\mathrm{E}_t\left[\exp\left(\frac{e_{i,t+1}\epsilon}{1+\epsilon\beta_k}\right)\right] \tag{C.21}$$

Since $e_{i,t+1}$ is normally distributed, $\exp\left(\frac{e_{i,t+1}\epsilon}{1+\epsilon\beta_k}\right)$ has a log-normal distribution. Let x denote $\frac{e_{i,t+1}\epsilon}{1+\epsilon\beta_k}$, then $x \sim N\left(0,\frac{\sigma_z^2\epsilon^2}{(1+\epsilon\beta_k)^2}\right)$ and $\exp(x) \sim LogNormal\left(0,\frac{\sigma_z^2\epsilon^2}{(1+\epsilon\beta_k)^2}\right)$. Use the fact that $\mathrm{E}[\exp(x)] = \exp(\mathrm{E}[x]+\frac{1}{2}\mathrm{Var}[x])$, so

$$\mathrm{E}_t\left[\exp\left(\frac{e_{i,t+1}\epsilon}{1+\epsilon\beta_k}\right)\right] = \exp\left(\frac{\sigma_z^2\epsilon^2}{2(1+\epsilon\beta_k)^2}\right) \tag{C.22}$$

Substitute (C.22) into (C.21) and take logs to get:

$$\ln\mathrm{E}_t\left[z_{i,t+1}^{\frac{\epsilon}{1+\epsilon\beta_k}}\right] = \frac{\rho\epsilon}{1+\epsilon\beta_k}\ln z_{i,t} + \frac{\sigma_z^2\epsilon^2}{2(1+\epsilon\beta_k)^2} \tag{C.23}$$

Finally, substitute (C.23) into (C.19) to get (4.13):

$$\begin{aligned}\ln k_{i,t}^U =& \epsilon\rho \ln z_{i,t} + (1+\epsilon\beta_k)\left\{\ln\left(\beta_k \beta_l^{\frac{\epsilon\beta_l}{1+\epsilon\beta_k}} \beta_m^{\frac{\epsilon\beta_m}{1+\epsilon\beta_k}}\right) - \ln(r_t+\delta)\right. \\ &\left. + \ln \mathrm{E}_t\left[\left(\frac{Z_{t+1}}{w_{t+1}^{\beta_l} p_{m,t+1}^{\beta_m}}\right)^{\frac{\epsilon}{1+\epsilon\beta_k}}\right] + \frac{\epsilon}{1+\epsilon\beta_k}\ln z_i + \frac{\sigma_z^2\epsilon^2}{2(1+\epsilon\beta_k)^2}\right\}\end{aligned} \tag{4.13}$$

C.1.4 Dispersion of MRPK within Unconstrained Firms

Using (4.13) and (C.12), the marginal revenue product of capital of an unconstrained firm is:

$$\begin{aligned}\ln \mathrm{MRPK}_{i,t}^U =& \frac{\epsilon}{1+\epsilon\beta_k}\ln(Z_t z_i z_{i,t}) - \frac{\epsilon\rho}{1+\epsilon\beta_k}\ln z_{i,t-1} - \ln\left(\beta_k \beta_l^{\frac{\epsilon\beta_l}{1+\epsilon\beta_k}} \beta_m^{\frac{\epsilon\beta_m}{1+\epsilon\beta_k}}\right) \\ & + \ln(r_{t-1}+\delta) - \ln \mathrm{E}_{t-1}\left[\left(\frac{Z_t}{w_t^{\beta_l} p_{m,t}^{\beta_m}}\right)^{\frac{\epsilon}{1+\epsilon\beta_k}}\right] - \frac{\epsilon}{1+\epsilon\beta_k}\ln z_i - \frac{\sigma_z^2\epsilon^2}{2(1+\epsilon\beta_k)^2} \\ & - \frac{\epsilon\beta_l}{1+\epsilon\beta_k}\ln w_t - \frac{\epsilon\beta_m}{1+\epsilon\beta_k}\ln p_{m,t} + \ln\left(\beta_k \beta_l^{\frac{\epsilon\beta_l}{1+\epsilon\beta_k}} \beta_m^{\frac{\epsilon\beta_m}{1+\epsilon\beta_k}}\right) \\ =& \frac{\epsilon}{1+\epsilon\beta_k}(\ln z_{i,t} - \rho\ln z_{i,t-1}) + \ln\left(\frac{Z_t}{w_t^{\beta_l} p_{m,t}^{\beta_m}}\right)^{\frac{\epsilon}{1+\epsilon\beta_k}} - \ln \mathrm{E}_{t-1}\left[\left(\frac{Z_t}{w_t^{\beta_l} p_{m,t}^{\beta_m}}\right)^{\frac{\epsilon}{1+\epsilon\beta_k}}\right] \\ & + \ln(r_{t-1}+\delta) - \frac{\sigma_z^2}{2(1-\beta_l-\beta_m)^2} \\ =& \frac{\epsilon}{1+\epsilon\beta_k} e_{i,t} + \ln\left(\frac{Z_t}{w_t^{\beta_l} p_{m,t}^{\beta_m}}\right)^{\frac{\epsilon}{1+\epsilon\beta_k}} - \ln \mathrm{E}_{t-1}\left[\left(\frac{Z_t}{w_t^{\beta_l} p_{m,t}^{\beta_m}}\right)^{\frac{\epsilon}{1+\epsilon\beta_k}}\right] + \ln(r_{t-1}+\delta) \\ & - \frac{\sigma_z^2\epsilon^2}{2(1+\epsilon\beta_k)^2}\end{aligned} \tag{C.24}$$

Assuming σ_z is the same for all firms, the dispersion of MRPK among unconstrained firms is:[1]

$$\mathrm{Var}_i(\ln \mathrm{MRPK}_{i,t}^U) = \psi_1 \mathrm{Var}_i(e_{i,t}) \tag{4.16}$$

where $\psi_1 \equiv \left(\frac{\epsilon}{1+\epsilon\beta_l}\right)^2$ and i denotes an unconstrained firm i. As can be seen, the dispersion of MRPK among unconstrained firms is only driven by the cross-section dispersion of the productivity innovation $e_{i,t}$.

C.1.5 Capital Demand of Financially Constrained Firms

Using the assumption on the financing sources of capital (4.9), when the borrowing constraint (4.10) is binding (i.e., $\lambda_{i,t} > 0$), firm i's capital demand $k_{i,t}^C$ is determined by its net worth:

$$k_{i,t}^C = \frac{n_{i,t}}{1-\phi(1-\delta)} \tag{C.25}$$

[1] Alternatively, $\mathrm{Var}_i(\ln \mathrm{MRPK}_{i,t}^U)$ can be found by substituting $\ln k_{i,t}^U$ in (4.13) for $\ln k_{i,t}$ in (4.15).

Taking logs yields:

$$\ln k_{i,t}^{C} = \ln n_{i,t} - \ln[1-\phi(1-\delta)] \tag{4.12}$$

It can be seen from the first order condition with respect to $k_{i,t}$ (C.2) that:

$$\lambda_{i,t} = \mathrm{E}_t\left[\beta_k \frac{p_{i,t+1}y_{i,t+1}}{k_{i,t}} - (r_t+\delta)\right] > 0 \tag{C.26}$$

which implies that the expected MRPK is greater than $(r_t+\delta)$. It can be shown that the investment of a constrained firm is determined by its cash flow $\mathrm{CF}_{i,t}$, which is the revenue net of wage payments, materials costs, and net interest payments on loans, i.e., $\mathrm{CF}_{i,t} \equiv p_{i,t}y_{i,t} - w_t l_{i,t} - p_{m,t}m_{i,t} - r_{t-1}b_{i,t-1}$, assuming that debt is not repaid in each period but rolled over.[2] Using the definitions for net worth (4.8) and cash flow, and $k_{i,t} = n_{i,t} + b_{i,t}$ (4.9),

$$\begin{aligned} n_{i,t+1} &\equiv p_{i,t+1}y_{i,t+1} - w_{t+1}l_{i,t+1} - p_{m,t+1}m_{i,t+1} - R_t b_{i,t} + (1-\delta)k_{i,t} \\ &= p_{i,t+1}y_{i,t+1} - w_{t+1}l_{i,t+1} - p_{m,t+1}m_{i,t+1} - r_t b_{i,t} - \delta k_{i,t} + n_{i,t} = \mathrm{CF}_{i,t+1} - \delta k_{i,t} + n_{i,t} \end{aligned} \tag{C.27}$$

where the firm's net income is equal to $p_{i,t+1}y_{i,t+1} - w_{t+1}l_{i,t+1} - p_{m,t+1}m_{i,t+1} - r_t b_{i,t} - \delta k_{i,t}$, and cash flow is the sum of net income and the depreciation of capital stock.

Using (C.27) and the binding collateral constraint (C.25),

$$k_{i,t}^{C} - k_{i,t-1}^{C} = \frac{1}{1-\phi(1-\delta)}(n_{i,t} - n_{i,t-1}) = \frac{1}{1-\phi(1-\delta)}(\mathrm{CF}_{i,t} - \delta k_{i,t-1}) \tag{C.28}$$

C.1.6 Dispersion of MRPK within Constrained Firms

Using (4.12) and (C.12), the marginal revenue product of capital of a constrained firm is:

$$\begin{aligned} \ln \mathrm{MRPK}_{i,t}^{C} =& \frac{\epsilon}{1+\epsilon\beta_k}\ln Z_{i,t} - \frac{1}{1+\epsilon\beta_k}\left[\ln n_{i,t-1} - \ln(1-\phi(1-\delta))\right] \\ &- \frac{\epsilon\beta_l}{1+\epsilon\beta_k}\ln w_t - \frac{\epsilon\beta_m}{1+\epsilon\beta_k}\ln p_{m,t} + \ln\left(\beta_k \beta_l^{\frac{\epsilon\beta_l}{1+\epsilon\beta_k}} \beta_m^{\frac{\epsilon\beta_m}{1+\epsilon\beta_k}}\right) \\ =& \frac{\epsilon}{1+\epsilon\beta_k}\ln Z_t + \frac{\epsilon}{1+\epsilon\beta_k}\ln z_{i,t} + \frac{\epsilon}{1+\epsilon\beta_k}\ln z_i - \frac{1}{1+\epsilon\beta_k}\ln n_{i,t-1} \\ &+ \frac{1}{1+\epsilon\beta_k}\ln[1-\phi(1-\delta)] - \frac{\epsilon\beta_l}{1+\epsilon\beta_k}\ln w_t - \frac{\epsilon\beta_m}{1+\epsilon\beta_k}\ln p_{m,t} \\ &+ \ln\left(\beta_k \beta_l^{\frac{\epsilon\beta_l}{1+\epsilon\beta_k}} \beta_m^{\frac{\epsilon\beta_m}{1+\epsilon\beta_k}}\right) \end{aligned} \tag{C.29}$$

where the last step uses $Z_{i,t} \equiv Z_t z_i z_{i,t}$. Using (C.29) and the AR(1) process for the idiosyncratic transitory productivity $z_{i,t}$ (4.5), the dispersion of MRPK among constrained

[2] Apart from the terminal period where the gross interests on debt $R_t b_t$ are repaid, assume in all the other periods, debt is rolled over and only net interests on debt $r_t b_t$ are repaid.

firms is given by:[3]

$$
\begin{aligned}
\text{Var}_i(\ln\text{MRPK}^C_{i,t}) =& \psi_1\text{Var}_i(\ln z_i) + \psi_1\text{Var}_i(e_{i,t}) + \psi_1\rho^2\text{Var}_i(\ln z_{i,t-1}) + \psi_2\text{Var}_i(\ln n_{i,t-1}) \\
& - \psi_3\text{Cov}_i(\ln z_i + \rho\ln z_{i,t-1}, \ln n_{i,t-1}) \\
=& \psi_1\text{Var}_i(e_{i,t}) + \text{Var}_i(\psi_1^{\frac{1}{2}}\ln z_i + \psi_1^{\frac{1}{2}}\rho\ln z_{i,t-1} - \psi_2^{\frac{1}{2}}\ln n_{i,t-1})
\end{aligned}
\tag{4.17}
$$

where $\psi_1 \equiv \left(\frac{\epsilon}{1+\epsilon\beta_k}\right)^2$, $\psi_2 \equiv \left(\frac{1}{1+\epsilon\beta_k}\right)^2$, $\psi_3 \equiv 2\frac{\epsilon}{(1+\epsilon\beta_k)^2}$, and i denotes a constrained firm i. Using (4.16) and (4.17), it can be seen that:

$$
\text{Var}_i(\ln\text{MRPK}^C_{i,t}) > \text{Var}_i(\ln\text{MRPK}^U_{i,t}) \tag{C.30}
$$

since $\text{Var}_i(\psi_1^{\frac{1}{2}}\ln z_i + \psi_1^{\frac{1}{2}}\rho\ln z_{i,t-1} - \psi_2^{\frac{1}{2}}\ln n_{i,t-1}) > 0$.

C.2 Decomposition of the Dispersion of MRPK

Suppose there are M_t firms in a given industry and N_t of them are unconstrained in a given period t, where $N_t \leqslant M_t$, and the remaining $M_t - N_t$ firms are constrained. The distribution of the observed lnMRPK in the data is a mixture of two distributions of $\ln\text{MRPK}^U$ (for unconstrained firms) and $\ln\text{MRPK}^C$ (for constrained firms). It is shown below that the variance of lnMRPK across all firms in a given industry and a given period t can be written in terms of the variances and means over the subgroups (unconstrained U and constrained C) of firms.

Let $X_{i,t}$ denote $\ln\text{MRPK}^U_{i,t}$ and $Y_{i,t}$ denote $\ln\text{MRPK}^C_{i,t}$ in a given time period. For simplicity, the subscripts i and t are suppressed for $X_{i,t}$ and $Y_{i,t}$ in the following proof. Order the firms in such a way that the first N_t firms according to the index i are unconstrained and the rest of firms are constrained (i.e., firms $N_t + 1$ to M_t). In a given

[3] Alternatively, $\text{Var}_i(\ln\text{MRPK}^C_{i,t})$ can be found by substituting $\ln k^C_{i,t}$ in (4.12) for $\ln k_{i,t}$ in (4.15).

industry-year:

$$
\begin{aligned}
&\mathrm{Var}_i(\mathrm{lnMRPK}_{i,t}) = \mathrm{E}_i(\mathrm{lnMRPK}_{i,t}^2) - \mathrm{E}_i(\mathrm{lnMRPK}_{i,t})^2 \\
=&\frac{1}{M_t}\left[\sum_{i=1}^{N_t}(\mathrm{lnMRPK}_{i,t}^U)^2 + \sum_{i=N_t+1}^{M_t}(\mathrm{lnMRPK}_{i,t}^C)^2\right] \\
&-\left[\frac{1}{M_t}\left(\sum_{i=1}^{N_t}\mathrm{lnMRPK}_{i,t}^U + \sum_{i=N_t+1}^{M_t}\mathrm{lnMRPK}_{i,t}^C\right)\right]^2 \\
=&\frac{1}{M_t}\left[\sum_{i=1}^{N_t}X^2 + \sum_{i=N_t+1}^{M_t}Y^2\right] - \left[\frac{1}{M_t}\left(\sum_{i=1}^{N_t}X + \sum_{i=N_t+1}^{M_t}Y\right)\right]^2 \\
=&\frac{N_t}{M_t}\frac{1}{N_t}\sum_{i=1}^{N_t}X^2 + \frac{M_t-N_t}{M_t}\frac{1}{M_t-N_t}\sum_{i=N_t+1}^{M_t}Y^2 - \left(\frac{N_t}{M_t}\frac{1}{N_t}\sum_{i=1}^{N_t}X + \frac{M_t-N_t}{M_t}\frac{1}{M_t-N_t}\sum_{i=N_t+1}^{M_t}Y\right)^2 \\
=&\frac{N_t}{M_t}\mathrm{E}_i(X^2) + \frac{M_t-N_t}{M_t}\mathrm{E}_i(Y^2) - \left(\frac{N_t}{M_t}\mathrm{E}_i(X) + \frac{M_t-N_t}{M_t}\mathrm{E}_i(Y)\right)^2 \\
=&\frac{N_t}{M_t}\mathrm{E}_i(X^2) + \frac{M_t-N_t}{M_t}\mathrm{E}_i(Y^2) - \left(\frac{N_t}{M_t}\right)^2\mathrm{E}_i(X)^2 - \frac{2N_t(M_t-N_t)}{M^2}\mathrm{E}_i(X)\mathrm{E}_i(Y) \\
&-\left(\frac{M_t-N_t}{M_t}\right)^2\mathrm{E}_i(Y)^2 \\
=&\frac{N_t}{M_t}\left[\mathrm{E}_i(X^2) - \mathrm{E}_i(X)^2\right] + \frac{N_t}{M_t}\left(1-\frac{N_t}{M_t}\right)\mathrm{E}_i(X)^2 + \frac{M_t-N_t}{M_t}\left[\mathrm{E}_i(Y^2) - \mathrm{E}_i(Y)^2\right] \\
&+\frac{M_t-N_t}{M_t}\left(1-\frac{M_t-N_t}{M_t}\right)\mathrm{E}_i(Y)^2 - \frac{2N_t(M_t-N_t)}{M^2}\mathrm{E}_i(X)\mathrm{E}_i(Y) \\
=&\frac{N_t}{M_t}\mathrm{Var}_i(X) + \frac{M_t-N_t}{M_t}\mathrm{Var}_i(Y) + \frac{N_t(M_t-N_t)}{M^2}[\mathrm{E}_i(X)^2 + \mathrm{E}_i(Y)^2 - 2\mathrm{E}_i(X)\mathrm{E}_i(Y)] \\
=&\frac{N_t}{M_t}\mathrm{Var}_i(X) + \frac{M_t-N_t}{M_t}\mathrm{Var}_i(Y) + \frac{N_t(M_t-N_t)}{M^2}[\mathrm{E}_i(X) - \mathrm{E}_i(Y)]^2
\end{aligned}
\tag{C.31}
$$

where $\mathrm{E}_i(X)$ denotes the mean of $\mathrm{lnMRPK}_{i,t}$ across all the unconstrained firms i and $\mathrm{E}_i(Y)$ denotes the mean of $\mathrm{lnMRPK}_{i,t}$ across all the constrained firms i in a given period t. Similarly, $\mathrm{E}_i(X^2)$ and $\mathrm{Var}_i(X)$ are defined over the subgroup of unconstrained firms and $\mathrm{E}_i(Y^2)$ and $\mathrm{Var}_i(Y)$ are defined over the subgroup of constrained firms.

The last term in (C.31) is the squared difference between the mean values of lnMRPK within the two subgroups of firms, weighted by the product of the two fractions $\frac{N_t}{M_t}$ and $\frac{M_t-N_t}{M_t}$, which are the proportions of unconstrained firms and constrained firms respectively for each industry-year.

C.3 Production Function Estimation

Once the revenue elasticities (β_k, β_l and β_m) for each two-digit NACE Rev.2 industry are estimated, the log of revenue-based productivity (TFPR) for firm i in a given industry at time t is the residual term after subtracting the weighted sum of inputs from $\ln(p_{i,t}y_{i,t})$:

$$\text{logTFPR}_{i,t} \equiv \ln Z_{i,t} = \ln(p_{i,t}y_{i,t}) - \beta_k \ln k_{i,t} - \beta_l \ln l_{i,t} - \beta_m \ln m_{i,t} \tag{C.32}$$

where $p_{i,t}y_{i,t}$ is measured by the nominal revenue, $k_{i,t}$, $m_{i,t}$ and $l_{i,t}$ are measured by the book value of fixed tangible assets, material costs, and the wage bill, respectively. Wage bill is used to measure $l_{i,t}$ to control for the quality differences of labor across firms, following Gopinath et al. (2017). Labor and materials are variable inputs, whereas capital $k_{i,t}$ is the state variable, which is equivalent to $k_{i,t-1}$ in the model described in Section 4.2.

This paper uses the Wooldridge (2009) estimation-based approach. Wooldridge (2009) show that the two-step estimation proposed by Olley and Pakes (1996) (OP) and Levinsohn and Petrin (2003) (LP) can be implemented in one step using GMM, by applying different instruments to each of the two equations. As he pointed out, there are two advantages of using the joint GMM estimation compared to the two-step methods. First, if the variable input (labor) is also determined by unobserved productivity and state variables, then the coefficient on labor is unidentified in the first-stage estimation (Ackerberg et al., 2006) and hence two-step estimation does not work in this case. Second, it is easy to obtain fully robust standard errors using joint estimation.

The capital, labor, and materials coefficients are estimated for each two-digit NACE Rev.2 industry separately. The use of two-digit industries is to make sure there are enough observations in each industry to carry out the estimation.[4] For each firm i within a two-digit industry in period t:

$$\ln(p_{i,t}y_{i,t}) = \beta_0 + \beta_k \ln k_{i,t} + \beta_l \ln l_{i,t} + \beta_m \ln m_{i,t} + \ln Z_{i,t} + \varsigma_{i,t} \tag{C.33}$$

where the sequence $\ln Z_{i,t}$ is the unobserved revenue-based productivity and $\varsigma_{i,t}$ is a sequence of shocks that are assumed to be conditional mean independent of current and past inputs. Under OP and LP, the unobserved productivity is proxied by an unknown function of capital and investment (under OP) or intermediate inputs (under LP):

$$\ln Z_{i,t} = f(\ln k_{i,t}, \ln m_{i,t}) \tag{C.34}$$

[4]In most countries, the manufacture of tobacco products (industry 12 from NACE Rev.2 code) appears to be quite concentrated and the number of firm-year observations is very small, so the revenue elasticities for this industry are not estimated in those countries. In this paper, at least 200 firm-year observations are required to implement this method.

where the log of material costs $\ln m_{i,t}$ is the proxy variable. To estimate β_k, β_l and β_m jointly, Wooldridge (2009) assumes:

$$\mathrm{E}_i(\varsigma_{i,t}|\ln l_{i,t}, \ln k_{i,t}, \ln m_{i,t}, \ln l_{i,t-1}, \ln k_{i,t-1}, \ln m_{i,t-1}, ..., \ln l_{i,1}, \ln k_{i,1}, \ln m_{i,1}) = 0 \quad \text{(C.35)}$$

where $t = 1, 2, ..., T$. It can be seen that serial dependence in the idiosyncratic shocks $\varsigma_{i,t}$ is allowed in the above assumption, since past values of $\varsigma_{i,t}$ do not appear in the conditioning set. The following sufficient condition is used to restrict the dynamics of the productivity process $\ln Z_{i,t}$:

$$\begin{aligned} &\mathrm{E}_i(\ln Z_{i,t}|\ln k_{i,t}, \ln l_{i,t-1}, \ln k_{i,t-1}, \ln m_{i,t-1}, ..., \ln l_{i,1}, \ln k_{i,1}, \ln m_{i,1}) \\ &= \mathrm{E}_i(\ln Z_{i,t}|\ln Z_{i,t-1}) \equiv g[f(\ln k_{i,t-1}, \ln m_{i,t-1})] \end{aligned} \quad \text{(C.36)}$$

where $g(.)$ is an unknown function representing the process of productivity $\ln Z_{i,t}$. The last equality follows from $\ln Z_{i,t-1} = f(\ln k_{i,t-1}, \ln m_{i,t-1})$. The above assumption means that $\ln k_{i,t}$, past outcomes on $(\ln l_{i,t}, \ln k_{i,t}, \ln m_{i,t})$, and all functions of these are uncorrelated with the innovations $e_{Z,i,t} = \ln Z_{i,t} - E(\ln Z_{i,t}|\ln Z_{i,t-1})$.

Using $\ln Z_{i,t} = f(\ln k_{i,t}, \ln m_{i,t})$ and $\ln Z_{i,t} = g[f(\ln k_{i,t-1}, \ln m_{i,t-1})] + e_{Z,i,t}$, the two equations used to identify β_k, β_l, and β_m are:

$$\ln(p_{i,t}y_{i,t}) = \beta_0 + \beta_k \ln k_{i,t} + \beta_l \ln l_{i,t} + \beta_m \ln m_{i,t} + f(\ln k_{i,t}, \ln m_{i,t}) + \varsigma_{i,t},\ \ t = 1, ..., T \quad \text{(C.37)}$$

$$\ln(p_{i,t}y_{i,t}) = \beta_0 + \beta_k \ln k_{i,t} + \beta_l \ln l_{i,t} + \beta_m \ln m_{i,t} + g[f(\ln k_{i,t-1}, \ln m_{i,t-1})] + \xi_{i,t},\ \ t = 2, ..., T \quad \text{(C.38)}$$

where $\xi_{i,t} \equiv e_{Z,i,t} + \varsigma_{i,t}$. The orthogonality condition on the error term for the first equation is (C.35), and for the second equation, it is:

$$\mathrm{E}_i(\ln\xi_{i,t}|\ln k_{i,t}, \ln l_{i,t-1}, \ln k_{i,t-1}, \ln m_{i,t-1}, ..., \ln l_{i,1}, \ln k_{i,1}, \ln m_{i,1}) = 0,\ \ t = 2, ..., T \quad \text{(C.39)}$$

These two different orthogonality conditions on the error terms for the two equations imply that different instruments can be used for each equation. For instance, the state variable (capital $\ln k_{i,t}$), any lagged inputs or functions of these variables can be used as instrumental variables for both equations. In addition, the intermediate inputs (investment or intermediate inputs $\ln m_{i,t}$) can also be used as instruments for the first equation.

I use the prodest (Rovigatti and Mollisi, 2016) in Stata to calculate the productivity measure used in this paper. A third-degree polynomial is used to estimate the unknown functions $f(.,.)$ and $g(.)$, as suggested by Petrin et al. (2004). $f(\ln k_{i,t}, \ln m_{i,t})$ is approximated by all polynomials of order three or less, (i.e., $(\ln k_{i,t})^{q_1}(\ln m_{i,t})^{q_2}$ where $q_1 + q_2 \leqslant 3$, with $q_1 \geqslant 0$ and $q_2 \geqslant 0$) and can be written as:

$$f(\ln k_{i,t}, \ln m_{i,t}) \approx \vartheta_0 + \mathbf{\Gamma}(\ln k_{i,t}, \ln m_{i,t})\boldsymbol{\vartheta} \equiv \vartheta_0 + \boldsymbol{\Gamma_{i,t}}\boldsymbol{\vartheta} \tag{C.40}$$

where $\mathbf{\Gamma}_{i,t} \equiv \mathbf{\Gamma}(\ln k_{i,t}, \ln m_{i,t})$ is a vector of $1 \times Q$ vector of functions (polynomials) and $\boldsymbol{\vartheta}$ is a vector of $Q \times 1$ parameters.[5] In addition, $g(.)$ is assumed to be approximated by a G-degree polynomial in $\ln Z_{i,t}$:

$$g(\ln Z_{i,t}) = \rho_0 + \rho_1 \ln Z_{i,t} + ...\rho_G(\ln Z_{i,t})^G \tag{C.41}$$

where $G = 1$ is used in the prodest package. Substitute the polynomial approximations for the unknown functions into (C.37) and (C.38) and rearrange to write the two equations as a vector of residuals $\boldsymbol{\Lambda_{i,t}(\theta)}$:

$$\boldsymbol{\Lambda_{i,t}(\theta)} = \begin{pmatrix} \varsigma_{i,t}(\boldsymbol{\theta}) \\ \xi_{i,t}(\boldsymbol{\theta}) \end{pmatrix}$$
$$= \begin{pmatrix} \ln(p_{i,t}y_{i,t}) - \alpha_0 - \beta_k \ln k_{i,t} - \beta_l \ln l_{i,t} - \beta_m \ln m_{i,t} - \boldsymbol{\Gamma_{i,t}}\boldsymbol{\vartheta} \\ \ln(p_{i,t}y_{i,t}) - \zeta_0 - \beta_k \ln k_{i,t} - \beta_l \ln l_{i,t} - \beta_m \ln m_{i,t} - \rho_1 \ln(\boldsymbol{\Gamma_{i,t-1}}\boldsymbol{\vartheta})... - \rho_G(\boldsymbol{\Gamma_{i,t-1}}\boldsymbol{\vartheta})^G \end{pmatrix} \tag{C.42}$$

where $\alpha_0 = \beta_0 + \vartheta_0$ and ζ_0 are the new intercepts and $\boldsymbol{\theta}$ is a vector of coefficients to be estimated. The assumption of exogenous instruments $\boldsymbol{\tau_{i,t}}$ gives rise to the following moment conditions:

$$\mathrm{E}_i[\boldsymbol{\tau'_{i,t}}\boldsymbol{\Lambda_{i,t}(\theta)}] = 0 \quad t = 2, ..., T \tag{C.43}$$

GMM estimation can then be applied to find the vector of coefficients $\hat{\boldsymbol{\theta}}$. All the instruments for the second equation are also valid for the first equation, while the first equation has two additional instruments, the contemporaneous values of $\ln l_{i,t}$ and $\ln m_{i,t}$. The instruments used are:

$$\boldsymbol{\tau_{i,t}} \equiv \begin{pmatrix} \boldsymbol{\tau_{i,t,1}} & 0 \\ 0 & \boldsymbol{\tau_{i,t,2}} \end{pmatrix}, \quad t = 2, ..., T \tag{C.44}$$

where

$$\boldsymbol{\tau_{i,t,1}} = (\ln l_{i,t}, \boldsymbol{\Gamma_{i,t}}) \tag{C.45}$$

$$\boldsymbol{\tau_{i,t,2}} = (\ln k_{i,t}, \ln l_{i,t-1}, \boldsymbol{\Gamma_{i,t-1}}) \tag{C.46}$$

where $\ln m_{i,t}$ is included in $\boldsymbol{\Gamma_{i,t}}$. A key difference in the sets of instruments is that $\boldsymbol{\tau_{i,t,2}}$ does not include the contemporaneous values of $\ln l_{i,t}$ and $\ln m_{i,t}$.

[5] Wooldridge (2009) assumes that $\boldsymbol{\Gamma_{i,t}}$ includes at least $\ln k_{i,t}$ and $\ln m_{i,t}$ separately to nest the linear version of $f(\ln k_{i,t}, \ln m_{i,t})$ as a special case.

Table C.1: Revenue Elasticities of Inputs and Correlations

	Revenue Elasticities			Correlations between $\Delta\ln\text{TFPR}_{i,t-1}$ and				
Country	β_k	β_l	β_m	$\Delta\ln\text{Sales}_{i,t-1}$	$\Delta\log\text{VA}_{i,t-1}$	$\Delta\ln\text{FTA}_{i,t}$	$\frac{\text{CF}_{i,t-1}}{\text{FTA}_{i,t-2}}$	$\Delta\ln n_{i,t-1}$
Bulgaria	0.09	0.31	0.44	0.595***	0.706***	0.039***	0.054***	0.170***
Croatia	0.03	0.28	0.67	0.441***	0.562***	0.030***	0.080***	0.171***
Czech Republic	0.02	0.34	0.52	0.537***	0.647***	0.031***	0.089***	0.213***
Finland	0.06	0.40	0.34	0.533***	0.570***	0.054***	0.127***	0.285***
France	0.04	0.49	0.29	0.530***	0.547***	0.054***	0.110***	0.285***
Germany	0.04	0.42	0.33	0.689***	0.410***	0.058***	0.066***	0.186***
Italy	0.06	0.32	0.40	0.525***	0.569***	0.046***	0.102***	0.158***
Korea	0.02	0.09	0.84	0.403***	0.634***	0.046***	0.086***	0.198***
Norway	0.02	0.39	0.34	0.577***	0.601***	0.047***	0.092***	0.233***
Poland	0.04	0.30	0.52	0.552***	0.651***	0.045***	0.079***	0.243***
Portugal	0.07	0.42	0.39	0.545***	0.416***	0.041***	0.152***	0.226***
Romania	0.14	0.29	0.45	0.560***	0.629***	0.047***	0.078***	0.209***
Serbia	0.11	0.26	0.56	0.504***	0.616***	0.031***	0.060***	0.092***
Slovakia	0.08	0.27	0.55	0.549***	0.633***	0.037***	0.070***	0.180***
Slovenia	0.06	0.36	0.42	0.476***	0.542***	0.031***	0.034***	0.112***
Spain	0.04	0.41	0.42	0.502***	0.358***	0.045***	0.115***	0.193***
Sweden	0.05	0.35	0.33	0.390***	0.545***	0.028***	0.090***	0.196***
Ukraine	0.11	0.46	0.38	0.579***	0.628***	0.025***	0.076***	0.115***
United Kingdom	0.04	0.22	0.60	0.491***	0.663***	0.060***	0.106***	0.226***

Note: The tables shows the mean revenue elasticities of capital, labor and materials (i.e., β_k, β_l, β_m) calculated using the Wooldridge (2009) estimation-based method, and the correlations between the estimated lagged productivity $\Delta\ln\text{TFPR}_{i,t-1}$ and different variables, including the lagged sales growth $\Delta\ln\text{Sales}_{i,t-1}$, lagged value added growth $\Delta\ln\text{VA}_{i,t-1}$, net capital investment or capital growth $\Delta\ln\text{FTA}_{i,t}$, lagged cash flow $\frac{\text{CF}_{i,t-1}}{\text{FTA}_{i,t-2}}$, and lagged net worth growth $\Delta\ln n_{i,t-1}$. The time period covered is early 1990s to 2015. The exact sample period differs across countries, as can be found in Table 4.1. The stars indicate the significance of the correlation coefficients. Note that Japan is excluded because material costs are not available to estimate TFPR.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C.4 Data Cleaning and Summary Statistics

The following cleaning steps are applied to datasets extracted from the Orbis Historical Financial database for each country:

- Drop if industry code is missing.
- Consolidation code: only keep C1, U1, U2
- Only keep the entire calendar year: drop if the number of months is not equal to 12
- Accounting year: following Kalemli-Ozcan et al. (2015), if the closing date is before June 1st, then it should be counted as the previous year.
- Basic reporting mistakes: 1) Drop if both operating revenue and sales are missing. 2) Drop negative number of employees, negative fixed tangible assets and negative sales. Note that operating revenue in Orbis equals the sum of sales, other operating revenues and stock variations, so operating revenue can be negative.
  3) Drop if interest paid, depreciation, long-term debt, short-term debt, employees cost and material costs are negative.
- Following Appendix A3 of Gopinath et al. (2017), drop if age is negative, where age is computed as the difference between year and incorporation year plus one.
- Keep only one filing type for each firm throughout the years. Each firm can have a mixture of two filing types throughout time, i.e., annual report and local registry filing (majority). I find that annual report is often associated with consolidated account (C1), whereas local registry filing is often associated with unconsolidated accounts (U1 or U2). Since empirical analysis looks at within-firm over time variation, it is important to make sure that each firm only has one filing type or consolidation code over time. Whenever a firm has a mixture of filing types across years, the filing type that has more observations is kept. If the two filing types occur with the same frequency for a given firm, then one filing type is chosen if it has greater availability of other variables.
- Keep either consolidated or unconsolidated account for each firm throughout the years. After the previous step, the consolidation code for a firm should be consistent over time.
- Drop duplicates: each firm can have multiple entries for the same year. Duplicates are dropped according to several criteria.
  1) Accounting years can differ across countries. The month of the closing date that has the largest observations is the preferred month. Suppose it is 12 (December), then when dropping duplicates based on month, then month 12 is kept if this also occurs most frequently within the firm over time, also conditional on firm id, year,

Ticker and industry code.

2) After the previous step, if there are still duplicates, drop the duplicate entry with missing Ticker, conditional on firm id, year, industry code, month of the account closing date, and total assets being the same.

3) After the first two steps, if there are still duplicates, drop the duplicate entry with missing ISIN number, conditional on firm id, year, industry code, month of the account closing date, Ticker and total assets being the same.

- The original dataset is in US dollars. Convert the variables (with monetary value) into domestic currency using the exchange rate variable in the dataset.

- This paper focuses on the manufacturing industry so that the capital stock can be well measured by the fixed tangible assets. For each country, only the manufacturing industry (two-digit NACE Rev.2 Code in the range of 10-33) is kept. The description for each two-digit industry can be found in Table C.5.

- Further cleaning: Missing operating revenue (used to calculate the sales growth) and missing or zero fixed tangible assets (used to measure capital stock $k_{i,t}$) are dropped.[6] Firm-year observations with fewer than 3 consecutive years are dropped, since in the empirical regressions, lagged growth rates are used. Years with fewer than 50 firms are dropped, which happens in the earlier sample period in some countries.

- Winsorization: before running regressions for each industry or country, variables are winsorized at the 1st and 99th percentiles in the relevant sample. Variables that need winsorization include: capital growth or firm investment, sales growth, value added growth, productivity growth, cash flow over lagged capital stock, net worth growth, net worth-to-assets ratio, cash-to-assets ratio. Variables such as log of MRPK and log of total assets do not have high kurtosis and winsorization is not necessary.

[6] In Orbis data, the variable 'operating revenue' represents the turnover or sales, while the 'sales' variable represents the net sales.

Table C.2: Summary Statistics of Selected Variables for Each Country in the Baseline Sample

	$\Delta \ln \text{FTA}_{i,t}$			$\Delta \ln \text{Sales}_{i,t}$			$\Delta \ln \text{VA}_{i,t}$			$\Delta \ln \text{TFPR}_{i,t}$		
Country	Mean	Median	sd	Mean	Median	sd	Mean	Median	sd	Mean	Median	sd
Bulgaria	0.059	-0.026	0.616	0.046	0.047	0.654	0.053	0.048	0.687	-0.009	-0.002	0.403
Croatia	0.010	-0.038	0.717	0.003	0.020	0.652	0.033	0.031	0.692	-0.009	-0.007	0.298
Czech Republic	0.049	-0.026	0.568	0.036	0.034	0.433	0.042	0.035	0.493	-0.002	-0.001	0.221
Finland	0.002	-0.066	0.421	0.017	0.026	0.398	0.027	0.033	0.366	0.005	0.007	0.208
France	-0.021	-0.102	0.527	0.028	0.024	0.216	0.029	0.026	0.208	0.003	0.004	0.138
Germany	0.041	-0.033	0.449	0.041	0.012	0.219	0.040	0.038	0.348	0.010	0.009	0.131
Italy	0.023	-0.033	0.505	-0.000	0.020	0.444	0.020	0.026	0.389	-0.007	0.001	0.210
Japan	0.008	-0.024	0.321	0.004	0.009	0.279	0.083	0.018	0.558			
Korea	0.154	-0.000	0.670	0.131	0.084	0.483	0.138	0.101	0.557	0.006	0.005	0.130
Norway	-0.002	-0.059	0.606	0.051	0.039	0.393	0.051	0.043	0.358	0.014	0.011	0.190
Poland	0.055	-0.019	0.462	0.045	0.048	0.389	0.052	0.052	0.402	0.003	0.003	0.199
Portugal	-0.030	-0.074	0.521	0.015	0.012	0.394	0.022	0.019	0.495	-0.004	0.001	0.225
Romania	0.086	-0.003	0.824	0.083	0.108	0.852	0.117	0.137	0.926	0.003	0.003	0.455
Serbia	0.135	-0.008	0.653	0.081	0.095	0.780	0.150	0.130	0.799	-0.002	-0.008	0.425
Slovakia	-0.009	-0.053	0.666	0.005	0.027	0.610	0.023	0.033	0.608	-0.007	-0.001	0.306
Slovenia	-0.057	-0.060	0.821	-0.018	0.034	0.706	0.045	0.034	0.440	-0.003	0.001	0.220
Spain	0.032	-0.038	0.496	0.011	0.027	0.333	0.022	0.034	0.401	-0.007	-0.000	0.187
Sweden	-0.034	-0.077	0.550	0.023	0.031	0.378	0.024	0.031	0.364	-0.006	0.007	0.303
Ukraine	0.023	-0.040	0.619	0.031	0.076	0.899	0.088	0.109	0.853	-0.015	-0.009	0.518
United Kingdom	-0.007	-0.039	0.442	0.028	0.030	0.316	0.052	0.050	0.372	0.005	0.005	0.125

Note: The table shows the mean, median and standard deviation for each of the four variables: capital investment or capital growth $\Delta \ln \text{FTA}_{i,t}$, sales growth $\Delta \ln \text{Sales}_{i,t}$, value added growth $\Delta \ln \text{VA}_{i,t}$, and the productivity growth $\Delta \ln \text{TFPR}_{i,t}$, where the productivity is estimated using the Wooldridge (2009) approach. The time period covered is early 1990s to 2015. The exact sample period differs across countries, as can be found in Table 4.1. Note that TFPR cannot be estimated for Japan due to the lack of data on material costs.

Table C.3: Summary Statistics of Selected Variables for Each Country in the Baseline Sample

	$\frac{\text{CF}_{i,t-1}}{\text{FTA}_{i,t-2}}$			$\Delta\ln n_{i,t-1}$			Net worth/Assets			Cash/Assets		
Country	Mean	Median	sd	Mean	Median	sd	Mean	Median	sd	Mean	Median	sd
Bulgaria	1.403	0.326	4.326	0.176	0.093	0.513	0.388	0.446	0.524	0.196	0.082	0.245
Croatia	1.266	0.313	4.540	0.093	0.046	0.508	0.260	0.269	0.495	0.087	0.032	0.132
Czech Republic	1.148	0.291	3.711	0.109	0.064	0.435	0.356	0.430	0.522	0.157	0.087	0.180
Finland	1.180	0.424	3.350	0.075	0.059	0.459	0.375	0.422	0.469	0.182	0.109	0.196
France	1.621	0.652	4.197	0.083	0.067	0.368	0.328	0.349	0.316	0.177	0.115	0.183
Germany	1.632	0.384	6.184	0.097	0.056	0.452	0.340	0.304	0.257	0.138	0.063	0.171
Italy	0.813	0.284	2.846	0.084	0.041	0.442	0.214	0.182	0.292	0.082	0.028	0.120
Japan	0.269	0.122	0.904	0.051	0.035	0.307	0.185	0.220	0.494	0.203	0.161	0.162
Korea	1.353	0.237	3.984	0.182	0.131	0.480	0.395	0.365	0.270	0.060	0.021	0.094
Norway	1.760	0.414	6.157	0.094	0.068	0.488	0.276	0.285	0.362	0.185	0.111	0.197
Poland	1.250	0.316	4.376	0.103	0.070	0.406	0.434	0.486	0.406	0.103	0.045	0.138
Portugal	0.434	0.254	2.685	0.089	0.053	0.438	0.216	0.268	0.504	0.130	0.058	0.170
Romania	1.289	0.231	5.066	0.193	0.085	0.944	0.047	0.214	0.980	0.112	0.041	0.165
Serbia	0.797	0.250	3.365	0.209	0.116	0.562	0.391	0.368	0.308	0.055	0.017	0.093
Slovakia	0.781	0.267	2.709	0.066	0.045	0.684	0.168	0.277	0.740	0.152	0.073	0.191
Slovenia	1.221	0.298	3.699	-0.011	0.044	0.783	0.478	0.488	0.337	0.104	0.039	0.152
Spain	0.617	0.258	2.210	0.101	0.067	0.366	0.284	0.283	0.363	0.116	0.060	0.142
Sweden	1.384	0.395	4.814	0.077	0.057	0.396	0.422	0.419	0.278	0.180	0.105	0.201
Ukraine	0.717	0.100	5.769	0.090	0.019	0.618	0.338	0.519	0.801	0.073	0.015	0.137
United Kingdom	2.018	0.432	7.571	0.087	0.071	0.529	0.246	0.348	0.657	0.149	0.069	0.194

Note: The table shows the mean, median and standard deviation for each of the four variables: lagged cash flow over twice lagged fixed tangible assets $\frac{\text{CF}_{i,t-1}}{\text{FTA}_{i,t-2}}$, lagged net worth growth $\Delta\ln n_{i,t-1}$, net-worth-to-assets ratio, and cash-to-assets ratio. The time period covered is early 1990s to 2015. The exact sample period differs across countries, as can be found in Table 4.1.

Table C.4: Correlations between Lagged Sales Growth and Different Variables for Each Country

Country	$\Delta\ln\text{FTA}_{i,t}$	$\frac{\text{CF}_{i,t-1}}{\text{FTA}_{i,t-2}}$	$\Delta\ln n_{i,t-1}$	$\Delta\ln\text{VA}_{i,t-1}$	$\Delta\ln\text{TFPR}_{i,t-1}$
Bulgaria	0.133***	0.187***	0.397***	0.815***	0.595***
Croatia	0.122***	0.166***	0.333***	0.727***	0.441***
Czech Republic	0.091***	0.156***	0.329***	0.759***	0.537***
Finland	0.102***	0.167***	0.322***	0.872***	0.533***
France	0.107***	0.169***	0.341***	0.892***	0.530***
Germany	0.091***	0.106***	0.198***	0.593***	0.689***
Italy	0.114***	0.181***	0.252***	0.846***	0.525***
Japan	0.075***	0.219***	0.302***	0.301***	
Korea	0.103***	0.175***	0.297***	0.729***	0.403***
Norway	0.080***	0.100***	0.247***	0.832***	0.577***
Poland	0.133***	0.155***	0.355***	0.842***	0.552***
Portugal	0.121***	0.182***	0.314***	0.659***	0.545***
Romania	0.069***	0.225***	0.501***	0.842***	0.560***
Serbia	0.166***	0.197***	0.346***	0.736***	0.504***
Slovakia	0.078***	0.154***	0.403***	0.807***	0.549***
Slovenia	0.023***	0.121***	0.732***	0.860***	0.476***
Spain	0.120***	0.168***	0.291***	0.607***	0.502***
Sweden	0.102***	0.160***	0.331***	0.803***	0.390***
Ukraine	0.114***	0.158***	0.272***	0.802***	0.579***
United Kingdom	0.105***	0.122***	0.262***	0.517***	0.491***

Note: The table shows the correlations between lagged sales growth $\Delta\ln\text{Sales}_{i,t-1}$ and different variables, including net capital investment or capital growth $\Delta\ln\text{FTA}_{i,t}$, lagged cash flow over twice lagged capital stock $\frac{\text{CF}_{i,t-1}}{\text{FTA}_{i,t-2}}$, lagged net worth growth $\Delta\ln n_{i,t-1}$, lagged value added growth $\Delta\ln\text{VA}_{i,t-1}$, and lagged productivity growth $\Delta\ln\text{TFPR}_{i,t-1}$, where the productivity is estimated using the Wooldridge (2009) approach. The stars indicate the significance of the correlation coefficients. Note that TFPR cannot be estimated for Japan due to the lack of data on material costs.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.5: Industry Classification: NACE Rev.2 Code for Manufacturing

Nace Code	Descriptions
10	Manufacture of food products
11	Manufacture of beverages
12	Manufacture of tobacco products
13	Manufacture of textiles
14	Manufacture of wearing apparel
15	Manufacture of leather and related products
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
17	Manufacture of paper and paper products
18	Printing and reproduction of recorded media
19	Manufacture of coke and refined petroleum products
20	Manufacture of chemicals and chemical products
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
22	Manufacture of rubber and plastic products
23	Manufacture of other non-metallic mineral products
24	Manufacture of basic metals
25	Manufacture of fabricated metal products, except machinery and equipment
26	Manufacture of computer, electronic and optical products
27	Manufacture of electrical equipment
28	Manufacture of machinery and equipment
29	Manufacture of motor vehicles, trailers and semi-trailers
30	Manufacture of other transport equipment
31	Manufacture of furniture
32	Other manufacturing
33	Repair and installation of machinery and equipment

Note: The table shows the NACE Rev. 2 Code for the two-digit industries in the manufacturing sector and their corresponding descriptions. More detailed industry classification can be found: https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF.

C.5 Exogenous Switching Regression Model

The likelihood function $L_{i,t}$ (4.31) of an observation in the exogenous switching regression model is derived below:

$$\begin{aligned} L_{i,t} =& f(\varepsilon_{C,i,t}|\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S})P(\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S}) \\ &+ f(\varepsilon_{U,i,t}|\varepsilon_{S,i,t} \leqslant -\boldsymbol{x_{S,i,t}\gamma^S})P(\varepsilon_{S,i,t} \leqslant -\boldsymbol{x_{S,i,t}\gamma^S}) \\ =& \frac{\int_{-\boldsymbol{x_{S,i,t}}\gamma^{\boldsymbol{S}}}^{\infty} f(\varepsilon_{C,i,t}, \varepsilon_{S,i,t})d\varepsilon_{S,i,t}}{P(\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S})} P(\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S}) \\ &+ \frac{\int_{-\infty}^{-\boldsymbol{x_{S,i,t}}\gamma^{\boldsymbol{S}}} f(\varepsilon_{U,i,t}, \varepsilon_{S,i,t})d\varepsilon_{S,i,t}}{P(\varepsilon_{S,i,t} \leqslant -\boldsymbol{x_{S,i,t}\gamma^S})} P(\varepsilon_{S,i,t} \leqslant -\boldsymbol{x_{S,i,t}\gamma^S}) \\ =& \int_{-\boldsymbol{x_{S,i,t}}\gamma^{\boldsymbol{S}}}^{\infty} f(\varepsilon_{C,i,t}, \varepsilon_{S,i,t})d\varepsilon_{S,i,t} + \int_{-\infty}^{-\boldsymbol{x_{S,i,t}}\gamma^{\boldsymbol{S}}} f(\varepsilon_{U,i,t}, \varepsilon_{S,i,t})d\varepsilon_{S,i,t} \\ =& f(\varepsilon_{C,i,t}) \int_{-\boldsymbol{x_{S,i,t}}\gamma^{\boldsymbol{S}}}^{\infty} f(\varepsilon_{S,i,t})d\varepsilon_{S,i,t} + f(\varepsilon_{U,i,t}) \int_{-\infty}^{-\boldsymbol{x_{S,i,t}}\gamma^{\boldsymbol{S}}} f(\varepsilon_{S,i,t})d\varepsilon_{S,i,t} \\ =& f(\varepsilon_{C,i,t})P(\varepsilon_{S,i,t} > -\boldsymbol{x_{S,i,t}\gamma^S}) + f(\varepsilon_{U,i,t})P(\varepsilon_{S,i,t} \leqslant -\boldsymbol{x_{S,i,t}\gamma^S}) \end{aligned} \tag{4.31}$$

where $f(\varepsilon_{C,i,t}|.)$ and $f(\varepsilon_{U,i,t}|.)$ denote general conditional probability densities and $f(.)$ is the marginal density. The fourth step uses the assumption that $\varepsilon_{C,i,t}$ and $\varepsilon_{U,i,t}$ are each independent from the error term $\varepsilon_{S,i,t}$ in the selection equation.

C.6 Robustness Checks

Table C.6: Capital Investment-Cash Flow Sensitivity and Marginal Revenue Product of Capital (MRPK)

Country	ΔlnTFPR	ΔlnTFPR $* d$	$\frac{CF}{FTA}$	$\frac{CF}{FTA} * d$	d(MRPK $> p70$)	Within R^2	Observations
Bulgaria	-0.000	0.037**	-0.001	0.024***	0.380***	0.0679	62,361
	(0.0059)	(0.0186)	(0.0022)	(0.0026)	(0.0117)		
Croatia	0.003	-0.020	0.005***	0.021***	0.420***	0.0682	76,801
	(0.0094)	(0.0333)	(0.0017)	(0.0021)	(0.0106)		
Czech Republic	0.003	0.005	0.006***	0.017***	0.409***	0.0686	106,834
	(0.0086)	(0.0277)	(0.0016)	(0.0020)	(0.0086)		
Finland	0.024***	0.009	0.010***	0.005***	0.287***	0.0559	107,782
	(0.0073)	(0.0200)	(0.0015)	(0.0018)	(0.0062)		
France	0.040***	0.007	0.021***	0.005***	0.302***	0.0712	972,611
	(0.0043)	(0.0118)	(0.0004)	(0.0005)	(0.0022)		
Germany	0.057***	-0.001	0.002	0.011***	0.285***	0.0552	60,566
	(0.0151)	(0.0370)	(0.0022)	(0.0025)	(0.0098)		
Italy	0.031***	0.007	0.018***	0.009***	0.293***	0.0512	1,198,195
	(0.0028)	(0.0072)	(0.0007)	(0.0008)	(0.0021)		
Korea	0.099***	-0.065**	0.004***	0.028***	0.574***	0.1074	341,295
	(0.0088)	(0.0318)	(0.0009)	(0.0012)	(0.0058)		
Norway	0.050***	-0.080**	0.010***	0.008***	0.374***	0.0669	73,268
	(0.0131)	(0.0405)	(0.0012)	(0.0014)	(0.0096)		
Poland	0.030***	0.000	0.008***	0.008***	0.280***	0.0505	83,927
	(0.0110)	(0.0293)	(0.0021)	(0.0023)	(0.0082)		
Portugal	0.033***	-0.001	0.007***	0.012***	0.304***	0.0430	252,792
	(0.0047)	(0.0157)	(0.0013)	(0.0016)	(0.0045)		
Romania	0.016***	0.066***	0.005***	0.013***	0.381***	0.0550	336,141
	(0.0032)	(0.0096)	(0.0006)	(0.0008)	(0.0047)		
Serbia	0.020***	0.025	0.007***	0.016***	0.384***	0.0642	99,047
	(0.0054)	(0.0154)	(0.0017)	(0.0021)	(0.0084)		
Slovakia	0.010	0.067*	0.005	0.026***	0.444***	0.0655	42,936
	(0.0126)	(0.0363)	(0.0035)	(0.0044)	(0.0147)		
Slovenia	0.026*	0.069	-0.000	0.045***	0.375***	0.0646	43,656
	(0.0148)	(0.0440)	(0.0047)	(0.0061)	(0.0144)		
Spain	0.007***	0.016*	0.013***	0.016***	0.280***	0.0520	960,187
	(0.0026)	(0.0097)	(0.0007)	(0.0009)	(0.0023)		
Sweden	0.016***	0.008	0.014***	0.004***	0.308***	0.0514	183,344
	(0.0052)	(0.0115)	(0.0009)	(0.0011)	(0.0058)		
Ukraine	0.012***	0.001	-0.001	0.011***	0.359***	0.0413	185,898
	(0.0031)	(0.0082)	(0.0013)	(0.0015)	(0.0064)		
United Kingdom	0.046***	0.002	0.021***	-0.003	0.194***	0.0350	94,157
	(0.0134)	(0.0330)	(0.0034)	(0.0036)	(0.0067)		

Note: The table shows the coefficients from regressing $\Delta\text{lnFTA}_{i,t}$ on lagged productivity growth $\Delta\text{lnTFPR}_{i,t-1}$ and lagged cash flow over twice lagged fixed tangible assets $\frac{CF_{i,t-1}}{FTA_{i,t-2}}$, and each of which interacted with a dummy that equals one if lagged log MRPK is in the top 30% and zero if otherwise. The last column shows the number of firm-year observations used in each regression. Firm and four-digit industry*year fixed effects are included in all regressions. Firm-level clustered standard errors are reported in parentheses.

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

Table C.7: Proportion of Constrained Firms for Each Two-digit Industry in Each Country

Country	10	13	14	16	18	20	22	23	25	26	27	28	31	32
Bulgaria	0.34	0.42	0.47	0.46	0.52	0.32	0.40	0.42	0.39	0.43	0.40	0.36	0.48	0.53
Croatia	0.40	0.44	0.41	0.41	0.40	0.47	0.41	0.43	0.41	0.47	0.48	0.42	0.43	0.39
Czech Republic	0.32	0.30	0.45	0.41	0.48	0.31	0.34	0.29	0.41	0.48	0.43	0.35	0.36	0.45
Finland	0.26	0.27	0.22	0.33	0.25	0.18	0.22	0.25	0.23	0.24	0.26	0.24	0.22	0.22
France	0.28	0.29	0.29	0.28	0.32	0.30	0.28	0.28	0.24	0.30	0.31	0.31	0.30	0.30
Germany	0.19	0.29	0.39	0.39	0.25	0.21	0.23	0.28	0.26	0.29	0.33	0.27	0.28	0.30
Italy	0.33	0.35	0.32	0.38	0.32	0.33	0.32	0.35	0.34	0.32	0.34	0.35	0.41	0.46
Japan	0.17	0.22	*	0.40	0.20	0.16	0.14	0.22	0.25	0.24	0.24	0.22	0.26	0.26
Korea	0.30	0.37	0.63	0.42	0.47	0.30	0.30	0.29	0.36	0.26	0.10	0.48	0.47	0.19
Norway	0.32	0.32	0.49	0.34	0.37	0.32	0.35	0.31	0.34	0.39	0.41	0.37	0.36	0.33
Poland	0.25	0.25	0.39	0.26	0.37	0.30	0.29	0.26	0.27	0.34	0.26	0.34	0.30	0.31
Portugal	0.28	0.32	0.30	0.32	0.33	0.26	0.32	0.26	0.37	0.32	0.39	0.40	0.33	0.30
Romania	0.36	0.37	0.19	0.19	0.37	0.36	0.35	0.12	0.44	0.34	0.18	0.35	0.36	0.15
Serbia	0.29	0.34	0.32	0.30	0.31	0.34	0.27	0.31	0.31	0.33	0.31	0.15	0.33	0.32
Slovakia	0.30	0.37	0.44		0.41	0.30	0.37	0.37	0.40	0.51	0.45	0.35	0.45	0.40
Slovenia	0.38	0.35	0.49	0.33	0.35	0.34	0.32	0.32	0.32	0.38	0.29	0.33	0.37	0.34
Spain	0.27	0.27	0.29	0.31	0.31	0.28	0.28	0.29	0.30	0.29	0.29	0.31	0.30	0.29
Sweden	0.32	0.41	0.39	0.29	0.32	0.35	0.33	0.29	0.32	0.29	0.36	0.36	0.39	0.30
Ukraine	0.30	0.37	0.30	0.32	0.29	0.31	0.31	0.31	0.33	0.34	0.34	0.35	0.30	0.33
United Kingdom	0.27	0.22	0.24	0.27	0.27	0.28	0.24	0.33	0.24	0.46	0.29	0.28	0.28	0.30
Unlisted - Listed	0.15	0.14	0.19	0.17	0.14	0.16	0.12	0.17	0.18	0.12	0.14	0.14	0.22	0.08
Number of Countries	11	9	9	3	4	10	7	11	11	12	11	12	7	7

Note: The table summarizes the results for the proportion of constrained firms from applying the switching regression model in the baseline analysis (where MRPK is computed using nominal revenue over fixed tangible assets) to 14 different two-digit industries selected based on the number of observations. Each column summarizes the average proportion of constrained firms across all firms and years in a given industry for different countries. The last two rows show the mean difference between the proportion of constrained firms in a sample of unlisted firms and that in a sample of listed firms, and the number of countries used to calculate this mean difference. The blank cell is due to the failure of convergence in log likelihood. * indicates none of MRPK, age or size is significant in the selection equation.

Table C.8: Credit Distortion for Each Two-digit Industry in Each Country

Country	10	13	14	16	18	20	22	23	25	26	27	28	31	32
Bulgaria	0.53	0.75	0.69	0.64	0.79	0.55	0.65	0.71	0.64	0.76	0.62	0.63	0.70	0.71
Croatia	0.61	0.72	0.60	0.62	0.67	0.73	0.61	0.65	0.64	0.67	0.69	0.61	0.67	0.58
Czech Republic	0.59	0.62	0.69	0.65	0.74	0.58	0.66	0.63	0.69	0.70	0.71	0.71	0.66	0.75
Finland	0.45	0.28	0.30	0.43	0.35	0.27	0.37	0.40	0.31	0.31	0.35	0.34	0.30	0.26
France	0.54	0.49	0.46	0.52	0.55	0.55	0.55	0.52	0.46	0.51	0.55	0.53	0.53	0.55
Germany	0.70	0.62	0.79	0.72	0.51	0.57	0.63	0.70	0.56	0.54	0.57	0.58	0.54	0.54
Italy	0.51	0.55	0.40	0.54	0.48	0.51	0.49	0.53	0.51	0.43	0.46	0.47	0.55	0.58
Japan	0.43	0.52	*	0.69	0.64	0.47	0.45	0.48	0.59	0.65	0.66	0.59	0.66	0.59
Korea	0.60	0.76	0.80	0.71	0.80	0.62	0.62	0.60	0.68	0.53	0.16	0.74	0.76	0.36
Norway	0.45	0.41	0.56	0.51	0.49	0.57	0.51	0.47	0.43	0.50	0.49	0.49	0.54	0.40
Poland	0.50	0.56	0.71	0.52	0.70	0.59	0.62	0.51	0.57	0.67	0.62	0.70	0.65	0.58
Portugal	0.53	0.61	0.54	0.61	0.68	0.57	0.57	0.54	0.68	0.51	0.70	0.68	0.64	0.51
Romania	0.47	0.52	0.30	0.31	0.53	0.48	0.50	0.23	0.57	0.51	0.29	0.50	0.49	0.25
Serbia	0.38	0.43	0.38	0.40	0.41	0.45	0.38	0.43	0.44	0.39	0.45	0.23	0.45	0.40
Slovakia	0.49	0.68	0.64		0.65	0.44	0.63	0.62	0.64	0.72	0.66	0.62	0.73	0.55
Slovenia	0.66	0.74	0.76	0.60	0.69	0.65	0.61	0.63	0.66	0.71	0.54	0.68	0.68	0.64
Spain	0.48	0.48	0.47	0.53	0.55	0.54	0.53	0.50	0.52	0.47	0.47	0.52	0.53	0.50
Sweden	0.53	0.62	0.56	0.50	0.50	0.53	0.59	0.54	0.56	0.40	0.55	0.56	0.67	0.43
Ukraine	0.40	0.47	0.42	0.43	0.47	0.39	0.44	0.44	0.47	0.41	0.45	0.46	0.42	0.42
United Kingdom	0.58	0.52	0.35	0.53	0.53	0.66	0.56	0.65	0.48	0.72	0.48	0.53	0.47	0.54
Unlisted - Listed	0.24	0.22	0.18	0.16	0.15	0.24	0.19	0.20	0.27	0.09	0.16	0.21	0.37	0.06
Number of Countries	11	9	9	3	4	10	7	11	11	12	11	12	7	7

Note: The table summarizes the results for credit distortion from applying the switching regression model in the baseline analysis (where MRPK is computed using nominal revenue over fixed tangible assets) to 14 different two-digit industries selected based on the number of observations. Each column summarizes the average credit distortion across all firms and years in a given industry for different countries. The last two rows show the mean difference between the credit distortion in a sample of unlisted firms and that in a sample of listed firms, and the number of countries used to calculate this mean difference. The blank cell is due to the failure of convergence in log likelihood. * indicates none of MRPK, age or size is significant in the selection equation.

Table C.9: Proportion of Constrained Firms for Each Two-digit Industry in Each Country

Country	10	13	14	16	18	20	22	23	25	26	27	28	31	32
Bulgaria	0.34	0.42	0.46	0.46	0.53	0.28	0.37	0.43	0.39	0.44	0.39	0.38	0.47	0.55
Croatia	0.40	0.45	0.40	0.40	0.40	0.48	0.41	0.42	0.41	0.48	0.46	0.42	0.43	0.40
Czech Republic	0.33	0.32	0.47	0.42	0.48	0.31	0.34	0.30	0.42	0.49	0.44	0.36	0.36	0.47
Finland	0.26	0.29	0.22	0.15	0.26	0.19	0.23	0.25	0.23	0.25	0.27	0.24	0.22	0.24
France	0.28	0.30	0.29	0.28	0.32	0.72	0.29	0.28	0.25	0.30	0.31	0.31	0.30	0.30
Germany	0.20		0.39	0.41	0.24	0.22	0.23	0.29	0.27	0.31	0.25	0.26	0.27	0.29
Italy	0.33	0.35	0.31	0.38	0.31	0.33	0.32	0.35	0.34	0.32	0.34	0.35	0.39	0.43
Japan	0.24	0.23	0.18	0.39	*	0.25	0.25	0.27	0.25	0.31	0.26	0.30	0.28	0.31
Korea	0.30	0.36	0.19	0.14	0.45	0.30	0.30	0.29	0.27	0.35	0.46	0.16	0.46	0.52
Norway	0.32	0.28	0.46	0.34	0.37	0.38	0.36	0.31	0.33	0.33	0.41	0.37	0.37	0.33
Poland	0.24	0.25	0.31	0.26	0.37	0.31	0.29	0.27	0.27	0.34	0.25	0.34	0.30	0.31
Portugal	0.29	0.32	0.29	0.32	0.32	0.31	0.32	0.35	0.36	0.32	0.38	0.39	0.34	0.30
Romania	0.36	0.36	0.36	0.35	0.24	0.51	0.19	0.19	0.41	0.35	0.36	0.36	0.41	0.23
Serbia	0.29	0.35	0.32	0.31	0.31	0.35	0.28	0.30	0.31	0.33	0.31	0.28	0.33	0.32
Slovakia	0.31	0.36	0.45	0.42	0.42	0.33	0.37	0.38	0.41	0.51	0.45	0.36	0.44	0.40
Slovenia	0.38	0.35	0.49	0.33	0.36	0.32	0.32	0.33	0.32	0.38	0.30	0.34	0.37	0.46
Spain	0.27	0.27	0.29	0.30	0.31	0.28	0.28	0.29	0.30	0.29	0.28	0.31	0.30	0.30
Sweden	0.32	0.38	0.39	0.30	0.32	0.41	0.33	0.30	0.32	0.30	0.39	0.37	0.40	0.30
Ukraine	0.28	0.39	0.30	0.30	0.28	0.33	0.30	0.30	0.32	0.33	0.32	0.36	0.29	0.35
United Kingdom	0.25	0.27	0.28	0.33	0.32	0.25	0.19	0.30	0.24	0.31	0.30	0.29	0.36	0.28
Unlisted - Listed	0.14	0.17	0.19	0.16	0.13	0.18	0.08	0.16	0.18	0.09	0.16	0.16	0.24	0.11
Number of Countries	10	6	7	2	3	9	5	9	9	10	9	11	4	5

Note: The table summarizes the results for the proportion of constrained firms from applying the switching regression model (where MRPK is computed using nominal value added over fixed tangible assets) to 14 different two-digit industries selected based on the number of observations. Each column summarizes the average proportion of constrained firms across all firms and years in a given industry for different countries. The last two rows show the mean difference between the proportion of constrained firms in a sample of unlisted firms and that in a sample of listed firms, and the number of countries used to calculate this mean difference. The blank cell is due to the failure of convergence in log likelihood. * indicates none of MRPK, age or size is significant in the selection equation.

Table C.10: Credit Distortion for Each Two-digit Industry in Each Country

Country	10	13	14	16	18	20	22	23	25	26	27	28	31	32
Bulgaria	0.58	0.74	0.69	0.67	0.80	0.47	0.64	0.73	0.64	0.77	0.67	0.68	0.71	0.76
Croatia	0.66	0.74	0.61	0.64	0.67	0.78	0.65	0.69	0.65	0.69	0.65	0.64	0.66	0.60
Czech Republic	0.66	0.67	0.70	0.69	0.72	0.62	0.67	0.64	0.70	0.72	0.74	0.70	0.68	0.78
Finland	0.45	0.42	0.31	0.23	0.35	0.29	0.39	0.41	0.33	0.31	0.36	0.36	0.33	0.27
France	0.56	0.53	0.49	0.54	0.56	0.89	0.57	0.56	0.48	0.53	0.58	0.55	0.55	0.55
Germany	0.72		0.82	0.73	0.48	0.59	0.64	0.68	0.56	0.52	0.55	0.56	0.51	0.54
Italy	0.58	0.57	0.41	0.58	0.49	0.55	0.52	0.57	0.53	0.45	0.49	0.50	0.55	0.57
Japan	0.47	0.53	0.43	0.69	*	0.42	0.50	0.40	0.52	0.62	0.71	0.63	0.57	0.60
Korea	0.62	0.74	0.23	0.27	0.78	0.61	0.61	0.59	0.60	0.63	0.76	0.28	0.77	0.79
Norway	0.57	0.42	0.57	0.56	0.56	0.72	0.56	0.55	0.47	0.53	0.55	0.56	0.61	0.50
Poland	0.59	0.57	0.62	0.55	0.72	0.61	0.62	0.54	0.59	0.71	0.61	0.71	0.69	0.57
Portugal	0.57	0.65	0.50	0.63	0.66	0.61	0.59	0.65	0.68	0.56	0.73	0.72	0.66	0.54
Romania	0.51	0.52	0.47	0.49	0.40	0.64	0.37	0.34	0.55	0.52	0.50	0.52	0.55	0.37
Serbia	0.41	0.45	0.38	0.42	0.45	0.51	0.40	0.39	0.44	0.42	0.46	0.42	0.49	0.42
Slovakia	0.55	0.63	0.68	0.64	0.71	0.58	0.66	0.68	0.66	0.76	0.68	0.64	0.75	0.58
Slovenia	0.71	0.77	0.76	0.64	0.71	0.67	0.68	0.66	0.70	0.72	0.57	0.71	0.71	0.78
Spain	0.53	0.52	0.47	0.56	0.54	0.56	0.55	0.52	0.54	0.49	0.48	0.55	0.53	0.52
Sweden	0.57	0.62	0.58	0.58	0.53	0.64	0.62	0.58	0.59	0.43	0.59	0.59	0.69	0.45
Ukraine	0.43	0.52	0.45	0.39	0.48	0.47	0.45	0.44	0.44	0.41	0.43	0.48	0.42	0.49
United Kingdom	0.56	0.54	0.51	0.61	0.63	0.56	0.41	0.58	0.50	0.53	0.50	0.58	0.62	0.50
Unlisted - Listed	0.21	0.29	0.26	0.22	0.18	0.21	0.09	0.26	0.29	0.12	0.24	0.24	0.45	0.19
Number of Countries	10	6	7	2	3	9	5	9	9	10	9	11	4	5

Note: The table summarizes the results for credit distortion from applying the switching regression model (where MRPK is computed using nominal value added over fixed tangible assets) to 14 different two-digit industries selected based on the number of observations. Each column summarizes the average credit distortion across all firms and years in a given industry for different countries. The last two rows show the mean difference between the credit distortion in a sample of unlisted firms and that in a sample of listed firms, and the number of countries used to calculate this mean difference. The blank cell is due to the failure of convergence in log likelihood. * indicates none of MRPK, age or size is significant in the selection equation.

Table C.11: Switching Regression Model of Firm Investment in Fabricated Metal Products Industry (Without Firm Fixed Effects)

	Unconstrained Regime		Constrained Regime				
Country	$\Delta \ln Sales_{i,t-1}$	$\frac{CF_{i,t-1}}{FTA_{i,t-2}}$	$\Delta \ln Sales_{i,t-1}$	$\frac{CF_{i,t-1}}{FTA_{i,t-2}}$	Observations	Prob > Chi2	df
Bulgaria	0.021***	0.007	0.109***	0.017***	4,243	0.0000	69
	(0.0061)	(0.0041)	(0.0324)	(0.0041)			
Croatia	0.031***	-0.004***	0.222***	0.016***	12,652	0.0000	69
	(0.0053)	(0.0016)	(0.0241)	(0.0026)			
Czech Republic	0.035***	-0.003**	0.150***	0.021***	25,421	0.0000	89
	(0.0040)	(0.0015)	(0.0207)	(0.0022)			
Finland	0.048***	-0.012***	0.120***	0.020***	27,429	0.0000	73
	(0.0037)	(0.0007)	(0.0232)	(0.0034)			
France	0.169***	-0.002***	0.358***	0.021***	170,850	0.0000	75
	(0.0045)	(0.0005)	(0.0177)	(0.0010)			
Germany	0.084***	-0.000	0.252***	0.008***	12,100	0.0000	91
	(0.0094)	(0.0008)	(0.0404)	(0.0023)			
Italy	0.032***	-0.012***	0.233***	0.028***	246,989	0.0000	87
	(0.0014)	(0.0005)	(0.0069)	(0.0011)			
Japan	0.026***	0.043**	0.138**	0.030	6,830	0.0000	79
	(0.0078)	(0.0172)	(0.0679)	(0.0243)			
Korea	0.010***	0.000	0.178***	0.024***	55,900	0.0000	53
	(0.0017)	(0.0007)	(0.0138)	(0.0022)			
Norway	0.044***	-0.003***	0.189***	0.018***	12,676	0.0000	71
	(0.0072)	(0.0006)	(0.0373)	(0.0023)			
Poland	0.068***	0.000	0.208***	0.011***	13,237	0.0000	75
	(0.0059)	(0.0014)	(0.0344)	(0.0025)			
Portugal	0.037***	-0.009***	0.272***	0.017***	47,373	0.0000	69
	(0.0035)	(0.0012)	(0.0162)	(0.0018)			
Romania	0.031***	-0.003***	0.102***	0.016***	44,863	0.0000	75
	(0.0023)	(0.0004)	(0.0088)	(0.0012)			
Serbia	0.028***	-0.000	0.149***	0.037***	12,866	0.0000	67
	(0.0033)	(0.0012)	(0.0207)	(0.0044)			
Slovakia	0.051***	-0.013***	0.124***	0.031***	10,806	0.0000	75
	(0.0063)	(0.0018)	(0.0267)	(0.0050)			
Slovenia	0.056***	-0.010***	0.315***	0.016***	12,476	0.0000	59
	(0.0065)	(0.0014)	(0.0335)	(0.0035)			
Spain	0.049***	-0.004***	0.270***	0.027***	193,141	0.0000	77
	(0.0019)	(0.0006)	(0.0128)	(0.0014)			
Sweden	0.082***	-0.009***	0.239***	0.016***	56,662	0.0000	71
	(0.0041)	(0.0005)	(0.0176)	(0.0015)			
Ukraine	0.015***	0.000	0.144***	0.009***	20,782	0.0000	63
	(0.0022)	(0.0007)	(0.0134)	(0.0018)			
United Kingdom	0.081***	-0.006***	0.237***	0.006***	26,117	0.0000	75
	(0.0057)	(0.0003)	(0.0289)	(0.0012)			

Note: The dependent variable is firm investment $\Delta \ln FTA_{i,t}$. The coefficients for lagged sales growth and lagged cash flow in two different investment regimes are reported. Four-digit industry and year fixed effects are included in the switching regression. The last two columns show the p-value for the likelihood ratio test and the degrees of freedom for the χ^2 distribution respectively. A small p-value suggests that the switching regression (less restrictive model) fits the data significantly better than an OLS regression. Robust standard errors are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

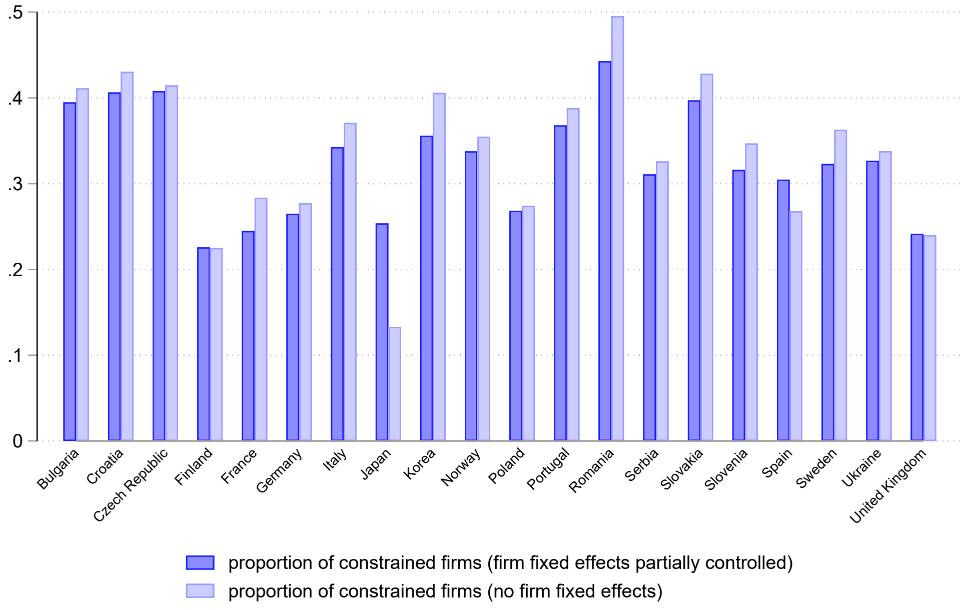
Table C.12: The Selection Equation of the Switching Regression in Fabricated Metal Products Industry (Without Firm Fixed Effects)

Country	Age	ln(Assets)	ln(MRPK)	$\frac{\text{Net worth}}{\text{Assets}}$	$\frac{\text{Cash}}{\text{Assets}}$	Fraction constrained
Bulgaria	-0.027***	-0.127***	0.790***	-0.449**	1.531***	0.41
	(0.0063)	(0.0417)	(0.0602)	(0.1784)	(0.3521)	
Croatia	-0.023***	-0.214***	0.907***	-0.504***	1.588***	0.43
	(0.0054)	(0.0229)	(0.0360)	(0.1102)	(0.3344)	
Czech Republic	-0.083***	-0.287***	0.933***	-0.308***	2.109***	0.41
	(0.0046)	(0.0190)	(0.0246)	(0.0710)	(0.1555)	
Finland	-0.020***	-0.142***	0.618***	0.041	0.398***	0.23
	(0.0022)	(0.0157)	(0.0225)	(0.0720)	(0.1246)	
France	-0.014***	-0.331***	1.032***	0.196***	1.186***	0.28
	(0.0008)	(0.0094)	(0.0122)	(0.0468)	(0.0615)	
Germany	-0.006***	-0.345***	0.789***	0.222	0.957***	0.28
	(0.0013)	(0.0257)	(0.0369)	(0.1709)	(0.2922)	
Italy	-0.012***	-0.236***	0.794***	-0.002	1.050***	0.37
	(0.0007)	(0.0064)	(0.0078)	(0.0371)	(0.0663)	
Japan	-0.024***	-0.142***	0.918***	-0.025	0.879*	0.13
	(0.0042)	(0.0432)	(0.0632)	(0.2414)	(0.4683)	
Korea	-0.029***	-0.267***	0.857***	-0.074	0.148	0.41
	(0.0021)	(0.0110)	(0.0138)	(0.0616)	(0.1640)	
Norway	-0.019***	-0.147***	0.671***	-0.129	0.402**	0.35
	(0.0049)	(0.0231)	(0.0273)	(0.1263)	(0.1662)	
Poland	-0.037***	-0.190***	0.712***	-0.345***	1.469***	0.27
	(0.0067)	(0.0244)	(0.0320)	(0.0962)	(0.2373)	
Portugal	-0.034***	-0.251***	1.108***	-0.106*	0.951***	0.39
	(0.0019)	(0.0156)	(0.0224)	(0.0616)	(0.1379)	
Romania	-0.039***	-0.149***	0.650***	-0.228***	1.038***	0.50
	(0.0034)	(0.0120)	(0.0181)	(0.0336)	(0.1250)	
Serbia	-0.010**	-0.059***	0.540***	-0.579***	1.758***	0.33
	(0.0042)	(0.0192)	(0.0276)	(0.1070)	(0.3700)	
Slovakia	-0.073***	-0.453***	0.858***	-0.381***	1.534***	0.43
	(0.0075)	(0.0294)	(0.0393)	(0.0943)	(0.2681)	
Slovenia	-0.043***	-0.333***	0.984***	-0.449***	1.930***	0.35
	(0.0054)	(0.0267)	(0.0393)	(0.1285)	(0.3262)	
Spain	-0.023***	-0.091***	0.887***	-0.044	0.740***	0.27
	(0.0011)	(0.0072)	(0.0089)	(0.0298)	(0.0634)	
Sweden	-0.012***	-0.320***	0.890***	0.310***	0.889***	0.36
	(0.0013)	(0.0131)	(0.0166)	(0.0743)	(0.0980)	
Ukraine	-0.016***	0.039***	0.494***	-0.246***	0.289	0.34
	(0.0034)	(0.0115)	(0.0165)	(0.0398)	(0.1909)	
United Kingdom	-0.004***	-0.113***	0.696***	-0.345***	-0.160	0.24
	(0.0012)	(0.0130)	(0.0241)	(0.0697)	(0.1396)	

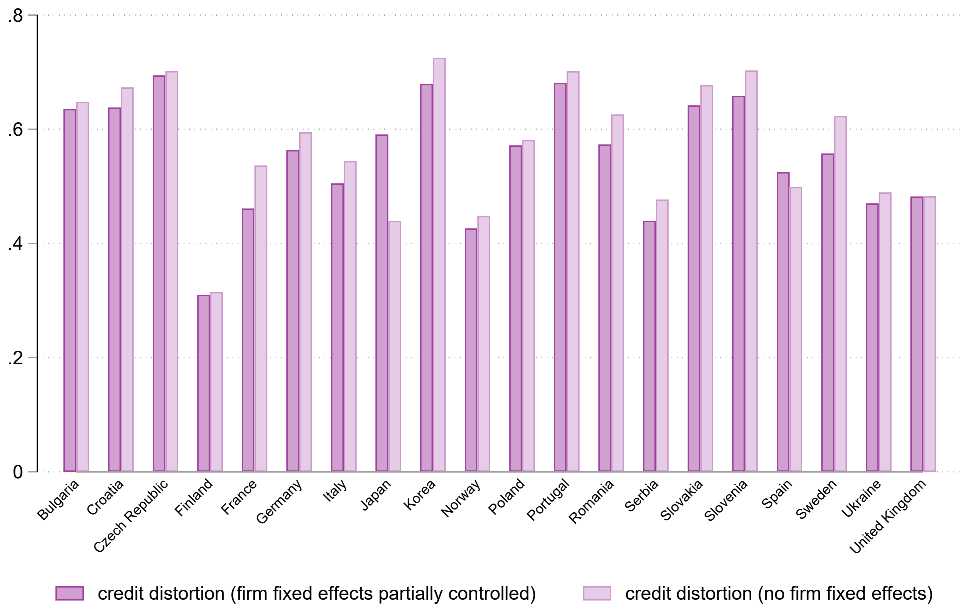
Note: The table shows the coefficients for the key variables in the selection equation that determines the probability of a firm being constrained, including age, log of assets, log of MRPK, net worth-to assets ratio, and cash-to-assets ratio, and the average proportion of constrained firms over the sample period. All variables apart from age are lagged. Four-digit industry and year fixed effects are included. The last column shows the average proportion of constrained firms over the sample period, where firms are classified as constrained based on the estimated posterior probabilities. Robust standard errors are reported in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure C.1: Proportion of Constrained Firms and Credit Distortion in Fabricated Metal Products Industry

(a) Proportion of Constrained Firms Using Different Fixed Effects



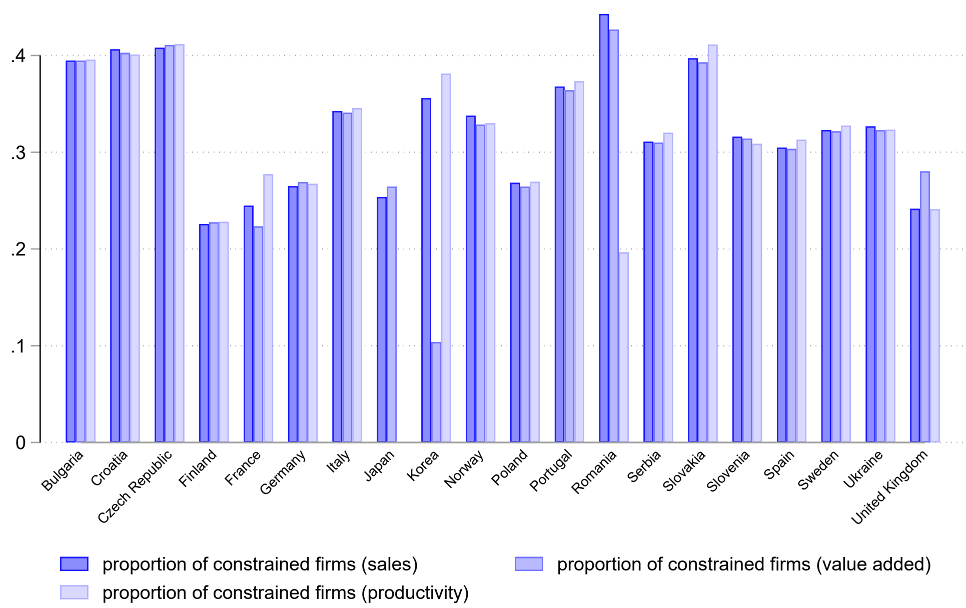
(b) Credit Distortion Using Different Fixed Effects



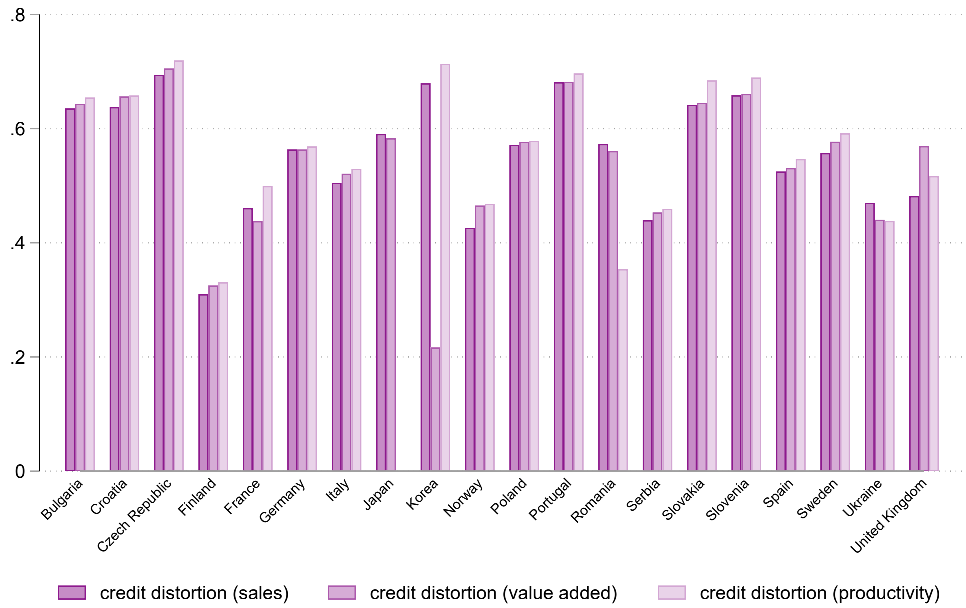
Note: In each graph, the corresponding measure is computed across all firms and years using the results from controlling for different fixed effects in the switching regression model: firm fixed effects partially controlled using the Hu and Schiantarelli (1998) approach and neglecting firm fixed effects. In both cases, four digit industry and year fixed effects are controlled. Graph (a) plots the fraction of constrained firms in industry 25 (manufacture of fabricated metal products) by NACE Rev.2 Code across 20 countries. Graph (b) plots credit distortion in percent points (i.e., the fraction of the observed dispersion (cross-section variance) of MRPK that is caused by the presence of constrained firms) in industry 25, which is computed based on (4.19). MRPK is computed as the nominal revenue divided by fixed tangible assets.
Data source: Orbis

Figure C.2: Proportion of Constrained Firms and Credit Distortion in Fabricated Metal Products Industry

(a) Proportion of Constrained Firms Using Different Proxies



(b) Credit Distortion Using Different Proxies



Note: In each graph, the corresponding measure is computed across all firms and years using the results from applying three different proxies for investment opportunity in the switching regression model: lagged sales growth, lagged value added growth and lagged productivity growth. Graph (a) plots the fraction of constrained firms in industry 25 (manufacture of fabricated metal products) by NACE Rev.2 Code across 20 countries. Graph (b) plots credit distortion in percent points (i.e., the fraction of the observed dispersion (cross-section variance) of MRPK that is caused by the presence of constrained firms) in industry 25, which is computed based on (4.19). MRPK is computed as the nominal revenue divided by fixed tangible assets.
Data source: Orbis

References

Ackerberg, D., Caves, K., and Frazer, G. (2006). Structural identification of production functions. Mimeo, UCLA Department of Economics.

Acosta Smith, J., Grill, M., and Lang, J. H. (2017). The leverage ratio, risk-taking and bank stability. ECB Working Paper No.2079.

Admati, A. R., DeMarzo, P. M., Hellwig, M. F., and Pfleiderer, P. C. (2013). Fallacies, irrelevant facts, and myths in the discussion of capital regulation: Why bank equity is not socially expensive. Stanford University Graduate School of Business Research Paper No.13-7.

Ağca, Ş. and Mozumdar, A. (2008). The impact of capital market imperfections on investment-cash flow sensitivity. *Journal of Banking & Finance*, 32(2):207–216.

Agénor, P.-R. and Montiel, P. J. (2015). *Development Macroeconomics*. Princeton University Press.

Allen, F., Carletti, E., and Marquez, R. (2011). Credit market competition and capital regulation. *The Review of Financial Studies*, 24(4):983–1018.

Allen, F. and Gale, D. (2000). *Comparing Financial Systems*. MIT Press. Chapter 8.

Allen, F. and Gale, D. (2004). Competition and financial stability. *Journal of Money, Credit, and Banking*, 36(3):453–480.

Almeida, H. and Campello, M. (2007). Financial constraints, asset tangibility, and corporate investment. *The Review of Financial Studies*, 20(5):1429–1460.

Alti, A. (2003). How sensitive is investment to cash flow when financing is frictionless? *The Journal of Finance*, 58(2):707–722.

Andrés, J. and Arce, O. (2012). Banking competition, housing prices and macroeconomic stability. *Economic Journal*, 122(565):1346–1372.

Anginer, D., Demirgüç-Kunt, A., and Zhu, M. (2014). How does competition affect bank systemic risk? *Journal of Financial Intermediation*, 23(1):1–26.

Ariss, R. T. (2010). On the implications of market power in banking: Evidence from developing countries. *Journal of Banking & Finance*, 34(4):765–775.

Asker, J., Collard-Wexler, A., and De Loecker, J. (2014). Dynamic inputs and resource (mis)allocation. *Journal of Political Economy*, 122(5):1013–1063.

Bai, Y., Lu, D., and Tian, X. (2018). Do financial frictions explain chinese firms' saving and misallocation? NBER Working Paper, No.24436.

Banerjee, A. V. and Duflo, E. (2005). Growth theory through the lens of development economics. *Handbook of Economic Growth*, 1:473–552.

Bartelsman, E., Haltiwanger, J., and Scarpetta, S. (2013). Cross-country differences in productivity: The role of allocation and selection. *The American Economic Review*, 103(1):305–334.

Beck, T., Colciago, A., and Pfajfar, D. (2014). The role of financial intermediaries in monetary policy transmission. *Journal of Economic Dynamics and Control*, 43:1–11.

Beck, T., De Jonghe, O., and Schepens, G. (2013). Bank competition and stability: Cross-country heterogeneity. *Journal of Financial Intermediation*, 22(2):218–244.

Beck, T., Demirgüç-Kunt, A., and Levine, R. (2006). Bank concentration, competition, and crises: First results. *Journal of Banking & Finance*, 30(5):1581–1603.

Beck, T., Demirgüç-Kunt, A., and Maksimovic, V. (2005). Financial and legal constraints to growth: Does firm size matter? *The Journal of Finance*, 60(1):137–177.

Berg, S. A. and Kim, M. (1998). Banks as multioutput oligopolies: An empirical evaluation of the retail and corporate banking markets. *Journal of Money, Credit and Banking*, 30(2):135–153.

Berger, A. N., Klapper, L. F., and Turk-Ariss, R. (2009). Bank competition and financial stability. *Journal of Financial Services Research*, 35(2):99–118.

Bernanke, B. and Gertler, M. (1989). Agency costs, net worth, and business fluctuations. *American Economic Review*, 79(1):14–31.

Bernanke, B., Gertler, M., and Gilchrist, S. (1996). The financial accelerator and the flight to quality. *Review of Economics and Statistics*, 78(1):1–15.

Bernanke, B. S., Gertler, M., and Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. *Handbook of Macroeconomics*, 1:1341–1393.

Besanko, D. and Thakor, A. V. (1992). Banking deregulation: Allocational consequences of relaxing entry barriers. *Journal of Banking & Finance*, 16(5):909–932.

Bikker, J. A. and Haaf, K. (2002). Competition, concentration and their relationship: An empirical analysis of the banking industry. *Journal of Banking & Finance*, 26(11):2191–2214.

Bikker, J. A., Shaffer, S., and Spierdijk, L. (2012). Assessing competition with the Panzar-Rosse model: The role of scale, costs, and equilibrium. *Review of Economics and Statistics*, 94(4):1025–1044.

Bolt, W. and Humphrey, D. (2015). A frontier measure of US banking competition. *European Journal of Operational Research*, 246(2):450–461.

Boyd, J. H. and De Nicolo, G. (2005). The theory of bank risk taking and competition revisited. *The Journal of Finance*, 60(3):1329–1343.

Busso, M., Madrigal, L., and Pagés, C. (2013). Productivity and resource misallocation in latin america. *The BE Journal of Macroeconomics*, 13(1):903–932.

Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398.

Caminal, R. and Matutes, C. (2002). Market power and banking failures. *International Journal of Industrial Organization*, 20(9):1341–1361.

Carbó, S., Humphrey, D., Maudos, J., and Molyneux, P. (2009). Cross-country comparisons of competition and pricing power in European banking. *Journal of International Money and Finance*, 28(1):115–134.

Carlson, M. A., Correia, S., and Luck, S. (2018). The effects of banking competition on growth and financial stability: Evidence from the national banking era. Available at SSRN:https://ssrn.com/abstract=3202489orhttp://dx.doi.org/10.2139/ssrn.3202489.

Carlstrom, C. T. and Fuerst, T. S. (1997). Agency costs, net worth, and business fluctuations: A computable general equilibrium analysis. *American Economic Review*, 87(5):893–910.

Carpenter, R. E. and Guariglia, A. (2008). Cash flow, investment, and investment opportunities: New tests using UK panel data. *Journal of Banking & Finance*, 32(9):1894–1906.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, 47(1):225–238.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45.

Claessens, S. and Laeven, L. (2004). What drives bank competition? Some international evidence. *Journal of Money, Credit and Banking*, 36(3):563–583.

Corbae, D. and D'Erasmo, P. (2011). A quantitative model of banking industry dynamics. manuscript, University of Wisconsin, Madison, and University of Maryland.

Corbae, D. and Levine, R. (2018). Competition, stability, and efficiency in financial markets. Mimeo. Available at http://online.wsj.com/public/resources/documents/corbae_levine_paper_0825.pdf?mod=article_inline.

Cuciniello, V. and Signoretti, F. M. (2015). Large banks, loan rate markup and monetary policy. *International Journal of Central Banking*, 11(3):141–177.

Cúrdia, V. and Woodford, M. (2015). Credit frictions and optimal monetary policy. NBER Working Paper, No.21820.

David, J. M. and Venkateswaran, V. (2017). The sources of capital misallocation. NBER Working Paper, No.23129.

De Bandt, O. and Davis, E. P. (2000). Competition, contestability and market structure in European banking sectors on the eve of EMU. *Journal of Banking & Finance*, 24(6):1045–1066.

Diamond, D. W. and Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91(3):401–419.

Dib, A. (2010). Banks, credit market frictions, and business cycles. Bank of Canada Working Paper, No.2010-24.

Dick, A. A. and Lehnert, A. (2010). Personal bankruptcy and credit market competition. *The Journal of Finance*, 65(2):655–686.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, 67(3):297–308.

Egan, M., Hortaçsu, A., and Matvos, G. (2017). Deposit competition and financial fragility: Evidence from the US banking sector. *American Economic Review*, 107(1):169–216.

Ehrmann, M., Gambacorta, L., Martínez-Pagés, J., Sevestre, P., and Worms, A. (2001). Financial systems and the role of banks in monetary policy transmission in the euro area. ECB Working Paper, No.105.

Faia, E., Laffitte, S., and Ottaviano, G. (2018). Foreign expansion, competition and bank risk. CEP Discussion Paper No.1567, Centre for Economic Performance, LSE.

Fazzari, S. M., Hubbard, R. G., Petersen, B. C., Blinder, A. S., and Poterba, J. M. (1988). Financing constraints and corporate investment. *Brookings Papers on Economic Activity*, 1988(1):141–206.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425.

Freixas, X. and Ma, K. (2015). Banking competition and stability: The role of leverage. Barcelona Graduate School of Economics, Working Papers 781.

Fu, X. M., Lin, Y. R., and Molyneux, P. (2014). Bank competition and financial stability in asia pacific. *Journal of Banking & Finance*, 38(1):64–77.

Gale, D. and Hellwig, M. (1985). Incentive-compatible debt contracts: The one-period problem. *The Review of Economic Studies*, 52(4):647–663.

Gambacorta, L. and Signoretti, F. M. (2014). Should monetary policy lean against the wind?: An analysis based on a DSGE model with banking. *Journal of Economic Dynamics and Control*, 43:146–174.

Garcia, R., Lusardi, A., and Ng, S. (1997). Excess sensitivity and asymmetries in consumption: An empirical investigation. *Journal of Money, Credit, and Banking*, 29(2):154–176.

Gerali, A., Neri, S., Sessa, L., and Signoretti, F. M. (2010). Credit and banking in a DSGE model of the euro area. *Journal of Money, Credit and Banking*, 42(s1):107–141.

Gertler, M. and Karadi, P. (2011). A model of unconventional monetary policy. *Journal of Monetary Economics*, 58(1):17–34.

Gertler, M., Kiyotaki, N., et al. (2010). Financial intermediation and credit policy in business cycle analysis. *Handbook of Monetary Economics*, 3(3):547–599.

Gertler, M., Kiyotaki, N., and Queralto, A. (2012). Financial crises, bank risk exposure and government financial policy. *Journal of Monetary Economics*, 59(supplement):S17–S34.

Gilchrist, S., Ortiz, A., and Zakrajsek, E. (2009). Credit risk and the macroeconomy: Evidence from an estimated DSGE model. Unpublished Manuscript, Boston University.

Gilchrist, S., Sim, J. W., and Zakrajšek, E. (2013). Misallocation and financial market frictions: Some direct evidence from the dispersion in borrowing cost. *Review of Economic Dynamics*, 16(1):159–176.

Goodfriend, M. and McCallum, B. T. (2007). Banking and interest rates in monetary policy analysis: A quantitative exploration. *Journal of Monetary Economics*, 54(5):1480–1507.

Gopinath, G., Kalemli-Özcan, Ş., Karabarbounis, L., and Villegas-Sanchez, C. (2017). Capital allocation and productivity in South Europe. *The Quarterly Journal of Economics*, 132(4):1915–1967.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal*, 7(1):98–119.

Hadlock, C. J. and Pierce, J. R. (2010). New evidence on measuring financial constraints: Moving beyond the KZ index. *The Review of Financial Studies*, 23(5):1909–1940.

Hafstead, M. and Smith, J. (2012). Financial shocks, bank intermediation, and monetary policy in a DSGE model. Unpublished Manuscript.

Hasan, I., Liu, L., and Zhang, G. (2016). The determinants of global bank credit-default-swap spreads. *Journal of Financial Services Research*, 50(3):275–309.

Hellmann, T. F., Murdock, K. C., and Stiglitz, J. E. (2000). Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? *American Economic Review*, 90(1):147–165.

Hovakimian, G. and Titman, S. (2006). Corporate investment with financial constraints: Sensitivity of investment to funds from voluntary asset sales. *Journal of Money, Credit, and Banking*, 38(2):357–374.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Hu, X. and Schiantarelli, F. (1998). Investment and capital market imperfections: A switching regression approach using US firm panel data. *Review of Economics and Statistics*, 80(3):466–479.

Hubbard, R. G., Kashyap, A. K., and Whited, T. M. (1995). Internal finance and firm investment. *Journal of Money, Credit and Banking*, 27(3):683–701.

Hull, J. (2012). *Options, Futures and Other Derivatives*. Upper Saddle River: Pearson Hall, 8th ed edition.

Hülsewig, O., Mayer, E., and Wollmershäuser, T. (2009). Bank behavior, incomplete interest rate pass-through, and the cost channel of monetary policy transmission. *Economic Modelling*, 26(6):1310–1327.

Iacoviello, M. (2005). House prices, borrowing constraints, and monetary policy in the business cycle. *American Economic Review*, 95(3):739–764.

Jiang, L., Levine, R., and Lin, C. (2017). Does competition affect bank risk? NBER Working Paper No.23080.

Jiménez, G., Lopez, J. A., and Saurina, J. (2013). How does competition affect bank risk-taking? *Journal of Financial Stability*, 9(2):185–195.

Kalemli-Ozcan, S., Sorensen, B., Villegas-Sanchez, C., Volosovych, V., and Yesiltas, S. (2015). How to construct nationally representative firm level data from the Orbis global database. NBER Working Paper No.21558.

Kaplan, S. N. and Zingales, L. (1997). Do investment-cash flow sensitivities provide useful measures of financing constraints? *The Quarterly Journal of Economics*, 112(1):169–215.

Keeley, M. C. (1990). Deposit insurance, risk, and market power in banking. *The American Economic Review*, 80(5):1183–1200.

Kiyotaki, N. and Moore, J. (1997). Credit cycles. *Journal of Political Economy*, 105(2):211–248.

Lamont, O., Polk, C., and Saaá-Requejo, J. (2001). Financial constraints and stock returns. *The Review of Financial Studies*, 14(2):529–554.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 70(2):317–341.

Liu, Z., Wang, P., and Zha, T. (2013). Land-price dynamics and macroeconomic fluctuations. *Econometrica*, 81(3):1147–1184.

Maddala, G. S. (1986). Disequilibrium, self-selection, and switching models. *Handbook of Econometrics*, 3:1633–1688.

Martinez-Miera, D. and Repullo, R. (2010). Does competition reduce the risk of bank failure? *The Review of Financial Studies*, 23(10):3638–3664.

Matutes, C. and Vives, X. (1996). Competition for deposits, fragility, and insurance. *Journal of Financial Intermediation*, 5(2):184–216.

Matutes, C. and Vives, X. (2000). Imperfect competition, risk taking, and regulation in banking. *European Economic Review*, 44(1):1–34.

Midrigan, V. and Xu, D. Y. (2014). Finance and misallocation: Evidence from plant-level data. *The American Economic Review*, 104(2):422–458.

Moshiriana, F., Nandab, V., Vadilyevc, A., and Zhanga, B. (2017). What drives investment–cash flow sensitivity around the world? An asset tangibility perspective. *Journal of Banking and Finance*, 77:1–17.

Mulier, K., Schoors, K., and Merlevede, B. (2016). Investment-cash flow sensitivity and financial constraints: Evidence from unquoted European SMEs. *Journal of Banking and Finance*, 73:182–197.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–97.

Oxenstierna, G. C. (1999). Testing for market power in the Swedish banking oligopoly. Stockholm University.

Perotti, E. C. and Suarez, J. (2002). Last bank standing: What do I gain if you fail? *European Economic Review*, 46(9):1599–1622.

Petrin, A., Poi, B. P., and Levinsohn, J. (2004). Production function estimation in Stata using inputs to control for unobservables. *Stata Journal*, 4(2):113–123.

Repullo, R. (2004). Capital requirements, market power, and risk-taking in banking. *Journal of Financial Intermediation*, 13(2):156–182.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707–720.

Restuccia, D. and Rogerson, R. (2013). Misallocation and productivity. *Review of Economic Dynamics*, 16(1):1–10.

Restuccia, D. and Rogerson, R. (2017). The causes and costs of misallocation. *Journal of Economic Perspectives*, 31(3):151–74.

Rotemberg, J. J. (1982). Monopolistic price adjustment and aggregate output. *The Review of Economic Studies*, 49(4):517–531.

Rovigatti, G. and Mollisi, V. (2016). PRODEST: Stata module for production function estimation based on the control function approach. Statistical Software Components S458239, Boston College Department of Economics.

Salas, V. and Saurina, J. (2003). Deregulation, market power and risk behaviour in Spanish banks. *European Economic Review*, 47(6):1061–1075.

Salop, S. C. (1979). Monopolistic competition with outside goods. *The Bell Journal of Economics*, 10(1):141–156.

Schaeck, K. and Cihák, M. (2007). Banking competition and capital ratios. IMF Working Paper 07/216.

Schaeck, K., Cihák, M., and Wolfe, S. (2009). Are competitive banking systems more stable? *Journal of Money, Credit and banking*, 41(4):711–734.

Schiantarelli, F. (1995). Financial constraints and investment: A critical review of methodological issues and international evidence. Boston College Working Paper, No.293.

Schiersch, A. and Schmidt-Ehmcke, J. (2010). Empiricism meets theory: Is the Boone-indicator applicable? DIW Berlin Discussion Paper, No. 1030.

Tabak, B. M., Fazio, D. M., and Cajueiro, D. O. (2012). The relationship between banking market competition and risk-taking: Do size and capitalization matter? *Journal of Banking & Finance*, 36(12):3366–3381.

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 21(2):265–293.

Uhde, A. and Heimeshoff, U. (2009). Consolidation in banking and financial stability in Europe: Empirical evidence. *Journal of Banking & Finance*, 33(7):1299–1311.

Vives, X. (2011). Competition policy in banking. *Oxford Review of Economic Policy*, 27(3):479–497.

Vives, X. (2016). *Competition and Stability in Banking*. Princeton University Press.

Whited, T. M. and Wu, G. (2006). Financial constraints risk. *The Review of Financial Studies*, 19(2):531–559.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, 104(3):112–114.

Wu, G. L. (2018). Capital misallocation in China: Financial frictions or policy distortions? *Journal of Development Economics*, 130:203–223.

Yeyati, E. L. and Micco, A. (2007). Concentration and foreign penetration in Latin American banking sectors: Impact on competition and risk. *Journal of Banking & Finance*, 31(6):1633–1647.