UCLA
UCLA Electronic Theses and Dissertations

Title
Banking, Financial Markets and the Implications of Financial Frictions on Firm Innovation and Growth

Permalink
https://escholarship.org/uc/item/6jj0z1dc

Author
Hu, Kun

Publication Date
2019

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
Los Angeles

Banking, Financial Markets and the Implications of Financial Frictions on Firm Innovation and Growth

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Kun Hu

2019

© Copyright by

Kun Hu

2019

ABSTRACT OF THE DISSERTATION

Banking, Financial Markets and
the Implications of Financial Frictions
on Firm Innovation and Growth

by

Kun Hu
Doctor of Philosophy in Economics
University of California, Los Angeles, 2019
Professor Lee Ohanian, Chair

My thesis consists of three chapters on banking, financial markets, financial frictions and their implications on firm innovation and growth.

Chapter 1:

Much of the research has focused on how financial constraints affect a firm's R&D expenditure, but little has been focused on the type of innovation. Using matched firm-patent panel data of U.S. public firms from 1997 to 2016, Chapter 1 shows that financially constrained firms switch toward internal (improve existing products) from external (develop new product lines) innovation. For an average firm, a one standard deviation decrease in cash-flow ratio leads to a 10-p.p. decrease in the percentage of external innovation. Also, the sensitivity of external innovation to cash flows is higher during the 2007-2009 financial crisis and for firms in sectors dependent on external financing. The empirical findings are robust to various panel regression specifications. Since external and internal innovations generate different quality improvement and growth potential, the interaction between financial constraints and the type of innovation provides a new mechanism by which finance affects firm and economic growth.

Chapter 2:

The firm size-growth relation plays a central role in many economic growth studies. Empirical literature shows that this relation depends on financial market conditions: sometimes small firms grow faster, but growth rates become independent of firm size

when frictions are high (Gibrat's law). In contrast to most growth models that assume a fixed size-growth relation, my framework allows it to vary with financial frictions. In the model, firms of different sizes grow by internal (improve existing products) and external (develop new products) R&D, but R&D expenditure is restricted by profits (or cash flows). This setup allows for rich interactions between size-growth relation and compositional change of R&D types in evaluating how financial frictions affect aggregate growth. The model is consistent with the fact that financially constrained firms switch to internal R&D and reduce R&D expenditure. The equilibrium is solved with neural networks. The estimated model suggests a significant drop in growth and welfare after the Great Recession. The importance of size-dependent policies that subsidize small firm R&D more than large firms is also quantified.

Chapter 3 (with TengTeng Xu and Udaibir S. Das):

We analyze how bank profitability impacts financial stability from both theoretical and empirical perspectives. We first develop a theoretical model of the relationship between bank profitability and financial stability by exploring the role of non-interest income and retail-oriented business models. We then conduct panel regression analysis to examine how the level and source of bank profitability affect risks for 431 publicly traded banks (U.S., advanced Europe, and GSIBs) from 2004 to 2017. Results reveal that profitability is negatively associated with both a bank's contribution to systemic risks and its idiosyncratic risks, and an over-reliance on non-interest income, wholesale funding and leverage is associated with higher risks. Low competition is associated with low idiosyncratic risks but high contribution to systemic risk. The paper's findings suggest that policy makers should strive to better understand the source of bank profitability, especially where there is an over-reliance on market-based non-interest income, leverage, and wholesale funding.

The dissertation of Kun Hu is approved.

Antonio E. Bernardo

Matthew Saki Bigio Luks

Pierre-Olivier Weill

Lee Ohanian, Committee Chair

University of California, Los Angeles

2019

To my parents,

who gave me—among so many other things—courage.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would first like to thank my committee chair, Prof. Lee Ohanian, who has given me invaluable guidance and support. Without his help and encouragement, this dissertation would not have been possible.

I would also like to extend my sincere gratitude to my committee members, Prof. Pierre-Olivier Weill, Prof. Saki Bigio and Prof. Antonio Bernardo, for their advice and assistance.

I am also grateful for the help and inspiration from my supervisors and colleagues at the International Monetary Fund (IMF), with special thanks to my co-authors TengTeng Xu and Udaibir S. Da of our working paper *Bank Profitability and Financial Stability*, as well as Martin Cihak, Javier Hamann and Hiroko Oura for their valuable feedback.

Last but not least, I would like to express my appreciation to all my classmates and friends at UCLA, for they have made my PhD study truly joyful and memorable.

VITA

2009 – 2013	B.S. (Economics and Finance), Hong Kong University of Science & Technology
2012 – 2012	Exchange Program, New York University
2013 – 2019	(Expected Graduation: 2019) PhD student (Economics), UCLA.
2014 – 2019	Teaching Assistant, Economics Department, UCLA. Taught intermediate macroeconomics/microeconomics/econometrics/finance/investment. Received the Teaching Assistant Award for excellence in teaching in 2016.
2018	Fund Internship Program, Monetary and Capital Market (MCM) Division, International Monetary Fund.

CHAPTER 1

Financial Constraints and External versus Internal Innovation

1.1 Introduction

The recovery from banking or other financial crises tends to be slow, usually accompanied by sluggish productivity growth. The lack of growth has reignited researchers' interest in how financial frictions affect firm innovation and growth. Much of the research has focused on the effect on firms' overall innovation efforts (e.g., total R&D expenditure), but less has examined whether firms will choose different types of innovation depending on the tightness of financial constraints. In this chapter, I will investigate the empirical relationship between financial constraints and type of innovation.

Different types of innovations generate heterogeneous quality improvements and growth potentials. Dependent on the mixtures of innovation projects, the economy can progress at various growth rates and feature distinct firm dynamics. If financial market conditions indeed influence firms' innovation decisions — external versus internal, radical versus incremental, then how financial constraints affect the type of innovation can shed new light on the implications of finance on firm and economic growth.

I follow [AK15] and categorize innovation into two types: external and internal innovation. A firm invests in the former to improve its existing products, and the latter to acquire new product lines. Innovation outcomes are measured by the patents in the data. The percentage of internal innovation is defined as the share of internal patents, which are those where 50% or more [1] of the given citations are to the prior inventions of the filing firm (i.e., self-citation). External patents are those with less than 50% self-citations.

[1] I allow for other thresholds different from 50% in the robustness check.

The dataset is constructed by matching firm and patent data from Compustat and PatentsViews. The sample consists of U.S. public firms from 1997 to 2016 (the decades before and after 2007). To measure a firm's tightness of financial constraints, I use a scaled cash-flow ratio, calculated as the product between cash flow to total asset ratio and the sector's median external financial dependence ratio (defined per [RZ98], it measures how much external financing is needed per dollar of capital expenditure). The first component measures the availability of internal finance, while the second component scales it by the likely scarcity, which reflects the importance of internal finance, as firms in sectors more dependent on external finance display higher sensitivity to internal funding availability.

I perform a correlated random effect IV Tobit panel regression to study the relationship between financial constraints and share of internal innovation. The results show that a one standard deviation decrease in cash flow ratio is associated with about a 10% increase in internal patent share, or about the median share. I also use the Great Recession as a natural experiment, and find that the sensitivity doubles in 2007–2009. The findings are robust to different variable definitions, identification methods, and econometric models.

In addition to documenting changes in the type of innovations, I find a decrease in overall innovation efforts during the Great Recession, measured by R&D expenditures, especially for small firms, which are more susceptible to financial constraints. The finding shows that small firms' (employment less than 100) real R&D expenditures drop by more than 20% in the Great Recession, three times more than those of large firms. This supports the argument that small firms were harder hit by the crisis[2].

My work is related to the empirical literature on the relation between R&D and financial markets[3]. [De 16] studies the effect of the Great Recession on R&D for U.S. firms and [Pei16] provides international evidence. Many papers also use natural experiments to analyze the effect of credit constraints on R&D: [CCW17] for the initiation of CDS trading; [NN14] concerning the Great Depression; [COS13, CMT15, ASZ13] for banking deregulation; and [Man18] with court rulings.

[2] Also see [Kab19] for evidence in the U.S., and [Sch17, Gar16] for Spain.

[3] See [HL10] and [KN15] for a survey.

1.2 Data

I employ the Compustat (firm data) and PatentsView (patents) dataset. The raw data spans Jan 1st, 1976 to May 16th, 2018. Firm and patent data are matched according to firm names. There are 85% exact matches and 15% fuzzy matches[4].

Following the selection method in [AK15], the main sample I use contains for-profit non-financial, non-farm, public U.S. firms that are continuously innovating. A firm is considered "continuously innovative" if it has conducted R&D or filed at least one patent in both of the two 5-year windows, i.e., 1997-2006 and 2007-2016, when it was operational. For patents, I consider only utility patents granted to U.S. corporations. In total, there are 31,895 firm-year observations, 3,187 firms and 569,304 utility patents in the 1997-2016 sample. Detailed data development process is included in Appendix 4.1

To get a sense of the sampled firms, consider a snapshot in 2015[5]. 57% of firms have more than 500 employees (53% in 1997) and 48% more than 1,000 (45% in 1997). In total, their R&D spending totaled $287 billion, accounting for 81% of total U.S. business R&D in 2015. Their total sales is $5,033 billion, or 56% of all U.S. for-profit, non-farm companies with more than five employees in 2015[6]. Other summary statistics are in Appendix 4.2.

1.3 R&D Expenditure Growth Rate During the Great Recession

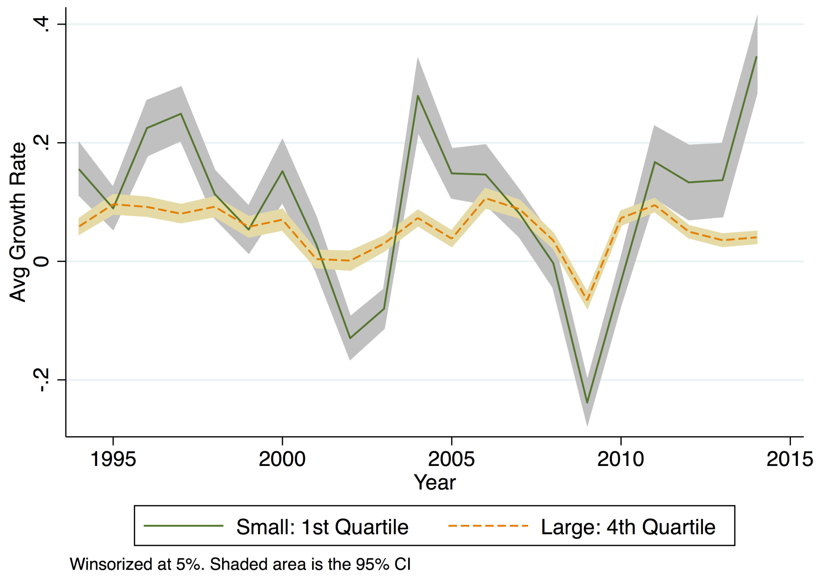
The empirical literature has argued that firms reduce the amount of R&D expenditure when financial constraints tighten. For example, [AAB12] documented the pro-cyclical R&D pattern for French firms with financial constraints. Since the link between financial constraints and R&D amount has been studied extensively in the literature, I will provide

[4]PatentsView is a public data source, which uses data derived from the US Patent and Trademark Office (USPTO) bulk data files. See `www.patentsview.org`. Due to typos or different abbreviation standards, firm names, even after standaraization, do not always match perfectly from different data sources. For fuzzy matching, I use the package FuzzyWuzzy in Python.

[5]2015 is the year of the most recent Business R&D and Innovation Survey (BRDIS) available as of the writing of this paper. See `www.nsf.gov/statistics/srvyindustry`

[6]The corresponding Compustat entry was REVT.

Figure 1.1: Real R&D Exp. Growth Rate by Firm Size



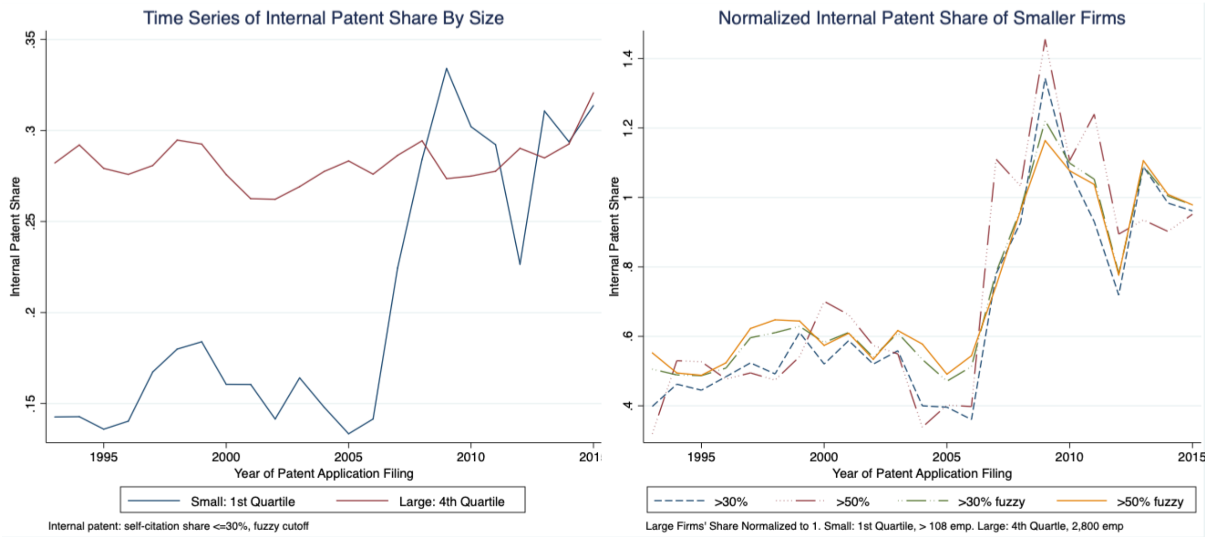
only some aggregate evidence in this section.

Figure 1.1 plots the change in the level of R&D for firms of different sizes (similar to [Bar07]). Real R&D expenditure growth rates drop more for smaller firms[7]. During the trough (2008-2009) of the crisis, for instance, large firms' R&D decreases by 7%, yet small firms slashes their R&D expenditure by more than 20%. Also visible from Figure 1.1 is that small firms' R&D spending is more volatile in general, suggesting that their capacity to smooth out their R&D is more limited.

1.4 Internal/External R&D and Financial Frictions

This section explores the relation between financial frictions and the type of R&D—internal vs. external innovation. I will first present aggregate-level evidence and then firm-level panel regressions.

Figure 1.2: Internal Innovation by Firm Size



1.4.1 Aggregate Evidence

I proxy internal R&D by internal patent share, which is defined based on patent citation patterns. More specificially, I first calculate the share of citations a patent made to its assignee's prior patents, termed self-citation[8], among its total backward citations. If a patent's self-citation share exceeds a certain threshold, it is counted as an internal patent. Internal patent share is the ratio of internal patents to total patents applied in a given year[9]. I use four thresholds to define internal patents: 30%, 50%, 30% (fuzzy) and 50% (fuzzy). All four assign a unit weight to a patent (i.e., count as one internal patent) when its self-citation share, x, passes the threshold, $A = 30\%$ or 50%, whereas the latter two *fuzzy* measures also assign fractional weight, x/A , when $x < A$[10].

Figure 1.2 plots the time series of internal patents by small and large firms, showing

[7]I categorize small and large firms according to the quartiles of firm employment. The quartile cut-offs are set based on the 2003 sample employment distribution: 108 and 2,800 employees respectively. Year 2003 is chosen arbitrarily and will not alter the results.

[8]For example, when a patent in iPhone X cites another Apple patent, this citation is called self-citation.

[9]For the aggregate measure plotted in Figure 1.2, I sum over all internal patents for small (large) firms by year, then divide by all filed patents of small (large) firms to arrive at the ratio. Alternatively, I can first calculate the internal patent share for each firm-year observation, and then take the average across firms in a given year. However, this measure of sample mean is biased, because not every firm filed patents every year [AK15].

[10]For instance, using a 50% fuzzy threshold, a patent with a 10% self-citation ratio is counted as 0.2 internal patents, but with a 50% threshold it is counted as 0 internal patent.

that small firms drastically increase their share of internal R&D after 2007. The left panel plots the internal patent share using a fuzzy 30% threshold. Before 2007, small firms' internal patent shares fluctuate around 15% and large around 30%, suggesting that small firms focus more on external innovation (consistent with [AK15]). After 2007, small firms' internal share doubles to 30% (or more), yet, by contrast, that of the large firms remains stable. This distinct response to financial crisis, a period with difficult credit conditions, will be closely reflected in the model, where small firms are more vulnerable to financial market imperfections.

The right panel serves as a robustness check. The conclusion from the left panel holds under various internal patent definitions. To make different time series comparable, I normalized large firms' internal patent shares to one. They all point out that small firms switch toward internal innovation during the Great Recession.

1.4.2 Firm-Level Evidence

To further analyze the effect of financial constraints on R&D type, I employ firm-level panel regression analysis. In particular, I estimate the following equation:

$$\begin{aligned}\text{Internal Patent Share}_{it} = \alpha_0 &+ \alpha_1 \text{EFD}_s \times \text{Cash Flow Ratio}_{i,t-1} \\ &+ \alpha_2 \text{EFD}_s \times \text{Cash Flow Ratio}_{i,t-1} \times \text{Crisis}_t \\ &+ \gamma' \text{Firm and Sector Controls}_{i,t-1} + u_i + v_t + \epsilon_{it}.\end{aligned} \tag{1.1}$$

Internal Patent Share$_{it}$ is a proxy for the internal innovation share. See section 1.4.1 for detailed definitions.

Cash Flow Ratio$_{i,t-1}$ denotes the lagged cash flow to total asset ratio[11].

Identification

The identification strategy relies on variation along two dimensions: (1) firms in sectors that are dependent on external finance vs. firms in other sectors; and (2) financial crisis (i.e., the Great Recession) vs. other time periods.

[11] Following [Bar07], I define cash flow = Income Before Extraordinary Items (Compustat item IB) + Depreciation and Amortization (Compustat item DP). I use cash flow to asset rather than to physical capital because R&D is considered intangible capital investment. Nonetheless, I included cash flow to net property, plant and equipment (PP&E) in the robustness check.

Across Firm Variation

To capture the sector-wise variation in combination with firm-level heterogeneity, I construct the interacted variable

$$\text{EFD}_s \times \text{Cash Flow Ratio}_{i,t-1}.$$

This variable measures the financial constraint faced by firm i, scaled by a sector-specific factor. On the one hand, cash flow ratio reflects the cash available to be deployed. One the other, EFD_s measures the importance of such cash for a typical firm in its sector. This is because EFD_s—*sector* (2-digit SIC code) external financial dependence ratio—is calculated as the sector median of *(capital expenditure - operational cash flow)/capital expenditure* using 1986-1992 Compustat data[12], so it is the external funding required as a fraction of capital expenditure. If we observe $\alpha_1 < 0$, it implies that as lower cash flow tightens constraints, firms perform more internal R&D.

The use of EFD_s comes with three advantages. First, estimated over a long period to smooth out cyclical fluctuation, EFD_s captures the financing needs of an industry, determined by structural characteristics. Second, based on data prior to the sample period, it is free from concurrent compounding factors. Third, because EFD_s is an industry-level statistics, it is unaffected by individual firms[13].

I employ four methods to address the endogeneity concern from $\text{Cash Flow Ratio}_{i,t-1}$. First, it is lagged by one period, which precludes the simultaneous feedback from firm innovation choice to financing constraints. Second, I include year and firm fixed effect u_i and v_t to account for any time- or firm-invariant unobservables (e.g., the availability of new ideas in a given year and a firm's innate tendency toward internal innovation). Third, multiple firm and sector controls are included for time and firm variant covariates. Last, cash flow terms are instrumented with the appropriated lagged cash flow ratios (lag three and further) to control for any remaining endogenous factors[14].

---

[12]External financial dependence follows [RZ98]. [AAB12] also show that innovative firms dependent on external financing are sensitive to credit conditions.

[13]See also in [LV13]

[14]Also note any omitted variables that are persistent (high autocorrelation) are controlled by firm fixed effect u_i. Transitory ones (low autocorrelation) are accounted by the lagged instruments and additional

Firm controls are Tobin's Q, firm sale growth rate and log-transformed variables, including total assets, capital expenditure, net PP&E, cash and short-term investment, short-term and long-term debt, net sales and the total number of patents owned by a firm (a.k.a. patent stock). Firm controls are lagged by one year to address endogeneity. Additionally, I add 2-digit SIC real value-added growth rates to reflect the industry wide demand factors.

Across Time Variation

The interaction of $\text{EFD}_s \times \text{Cash Flow Ratio}_{i,t-1} \times \text{Crisis}_t$ captures the potential change in the impact of financing constraints during the Great Recession, essentially treating the crisis as a natural experiment. Crisis_t is a binary variable, equaling 1 if $t = 2007, 2008$ or 2009. If the crisis exacerbated the financing constraints *even if* cash flow and external financial dependence stayed the same, we would expect $\alpha_2 < 0$. A significant and negative α_2 thus signals deteriorated financial conditions that are not explained by a firm's own characteristics.

Econometric Assumptions

However, there are some complications arising from the "small N large T " panel data structure and the fact that the dependent variable is left and right censored. Next, I explain my econometric assumptions to cope with these difficulties.

In the preferred specification, I use a correlated random effect Tobit model with instrument variables.

The internal patent share is left-censored at 0 (1/3 to 1/2 of the observations) and right-censored at 100% (1% of the observations), because firms with 0 or 100% internal patent share can still have different degrees of "internalness" as I do not control for the actual contents of patents: some may conduct more exploratative R&D than others. Thus, it is more appropriate to use a Tobit regression.

A concern for a nonlinear regression like Tobit is that one cannot estimate u_i using fixed-effect models, due to the "incidental parameter bias." I follow the Chamberlain-Mundlak approach of correlated random effect by assuming

$$u_i = a_0 + a_1'\bar{x}_i + a_2'\bar{z}_i + e_i$$

control variables.

where $\bar{x_i} = \sum x_{it}/T$, $\bar{z_i} = \sum z_{it}/T$, x_i is the independent variable(s), z_i other controls, and T the number of years when a firm is operational.

To further overcome potential endogenity issues, I used the 3rd to 8th lags of independent variables as instruments[15].

Regression Results

The results are shown in Table 1.1. Following [AK15], I use 50% as the threshold of internal patent definition. The preferred specification, IV Tobit with correlated random effects, is presented in Column (5). For a robustness check, I also adopted a fixed effect static panel regression (Model 1), dynamic panel system-GMM estimator developed by [AB95] and [BB98] (Model 2), random effect Tobit regression (Model 3) and correlated random effect Tobit without instrument variables (Model 4).

The first coefficient, or α_1 in (1.1), is negative and statistically significant, which implies that when financial constraints loosen due to the increased internal funding (i.e., cash flow ratio increases), firms lean toward more external (exploratory) innovation (and thus internal patent ratio decreases), especially firms in sectors more dependent on external financing (higher EFD_s). The coefficients are also economically significant. A one standard deviation decrease in the independent variable (0.85) is associated with a 9.3 percentage point increase in internal patent share, more than the median share of 5%.

The estimates of α_2 in (1.1) are presented next. They are mostly negative and significant as well, except for the regression using a 50% fuzzy threshold where the estimates are insignificant at a 10% confidence level. In terms of magnitude, they are close to that of α_1, meaning that the relation between cash flow and R&D becomes much more significant during the 2007-2009 financial crisis. Everything else equal, firms are more sensitive to internal funding and more susceptible to credit market conditions. One interpretation is that firms become much more risk averse and thus avert external innovation. Alternatively, firms find it harder to secure outside funding from banks or financial markets and thus are more cautious in their R&D projects. Either interpretation suggests an adverse shock to firm R&D decisions.

[15] The construction of instrument variables resembles Arellano-Bond. See [Roo09]

Table 1.1: Regression of Internal Patent Share and Financial Constraints

		(1)	(2)	(3)	(4)	(5)
Models		Fixed Effect	System GMM	RE Tobit	CRE Tobit	IV Tobit
Dependent Variable:						
Internal patent share, defined as patents with a self-citation share larger than or equal to 50%						
Independent Variable:						
(Lag) Cash Flow/Asset $\times$ (1986-1992) Sector External Financial Dependence						
α_1	...	-0.00220	-0.0191**	-0.0358**	-0.0181*	-0.109***
		(0.0100)	(0.00847)	(0.0181)	(0.0109)	(0.0421)
α_2	$\times$Crisis	-0.0562**	-0.0728***	-0.118***	-0.105***	-0.0886*
		(0.0261)	(0.0166)	(0.0349)	(0.0351)	(0.0519)
Observations		6,239	5,300	6,239	6,217	6,217
Number of firms		1,365	1,084	1,365	1,365	1,343
Controls and Year FE		Y	Y	Y	Y	Y
Firm FE		Y	Y	Random	Correlated	Correlated
Instrument variables			Y			Y

a. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Sample: 2002-2016. All variables are winsorized at 1% and 99%. Model (1) uses a within estimator with std. err. clustered by firm. Model (2) employs the Arellano-Bover/Blundell-Bond two-step estimator with the 3rd till 8th lags of indep. variables as instruments. Model (3) is a random effect Tobit model. Model (4) applies a correlated random effect Tobit model, a.k.a. Chamberlain-Mundlak approach, where unobserved time-invariant effect is estimated as a function of sample-period averages of controls and independent variables (std. err. clustered by firm). Model (5) uses lagged indep. variables (3rd and deeper) as instruments on top of Model (4), and adopts Newey's (1987) two-step estimator. **Self-citation**: when a patent cites its assignee's prior patents. Self-citation share for a patent = self-citations/total citations made. **External financial dependence** follows [RZ98] using 1986-1992 Compustat data, defined as the sector (2-digit SIC) median of (capital expenditure - operational cash flow) / capital expenditure. **Firm controls** are lagged by one year, including Tobin's Q and log-transformed total assets, capital expenditure, net PP&E, cash and short-term investment, short-term and long-term debt, net sales, sale growth rate and patent stock. **Sector controls** are real value-added growth rates. For (2), the p-value of Hansen and autocorrelation test is 0.163 and 0.382 (valid IVs). For (5), the Wald exogeneity test p-value is 0.0457 (justify using IVs).

1.5 Robustness Check

I performed various robustness checks with different samples, variable definitions and identification methods. All robustness check regressions are based on an IV correlated random effect Tobit (Model 5 in Table 1.1).

The first set of regressions adopt different definitions of internal patents, shown in Table 1.2. As evident in the table, both estimates of α_1 and α_2 remain negative and significant.

The second set of results employ the 50% threshold and econometric assumptions set forth in Model (5) in Table 1.1. The results are listed in Table 1.3. In Column (1), I include the full set of interaction terms[16] among EFD_s, Crisis_t and Cash Flow $\text{Ratio}_{i,t-1}$. The regression equation is now

$$\begin{aligned}\text{Internal Patent Share}_{it} =& \alpha_0 + \alpha_1 \text{EFD}_s \times \text{Cash Flow Ratio}_{i,t-1} \\ &+ \alpha_2 \text{EFD}_s \times \text{Cash Flow Ratio}_{i,t-1} \times \text{Crisis}_t \\ &+ \alpha_3 \text{Cash Flow Ratio}_{i,t-1} \times \text{Crisis}_t + \alpha_4 \text{Cash Flow Ratio}_{i,t-1} \\ &+ \gamma' \text{Firm and Sector Controls}_{i,t-1} + u_i + v_t + \epsilon_{it}.\end{aligned}$$

Both α_1 and α_2 show conclusions similar to those in Table 1.1.

I expand the sample by another ten years (prior to 2002) to be 1992-2016, shown in Column (2).

For Column (3), I replace the continuous variable EFD_s by a binary variable IEFD_s. $\text{IEFD}_s = 1$ if $\text{EFD}_s >$ median EFD_s across all industries. Firms with $\text{IEFD}_s = 1$ are considered the treatment group, and Crisis_t is a natural experiment. I then adopt regression discontinuity by estimating

$$\begin{aligned}\text{Internal Patent Share}_{it} =& \alpha_0 + \alpha_1 \text{IEFD}_s \times \text{Cash Flow Ratio}_{i,t-1} \\ &+ \alpha_2 \text{IEFD}_s \times \text{Crisis}_t \times \text{Cash Flow Ratio}_{i,t-1} \\ &+ \alpha_3 \text{Cash Flow Ratio}_{i,t-1} \times \text{Crisis}_t + \alpha_4 \text{Cash Flow Ratio}_{i,t-1} \\ &+ \gamma' \text{Firm and Sector Controls}_{i,t-1} + u_i + v_t + \epsilon_{it}.\end{aligned}$$

[16]Common in the literature using external financial dependence, the full interaction set is not necessary. See [RZ98, LV13]. Note that here it is not necessary to include the term EFD_s separately, as the sector time-invariant effect is already captured by the firm fixed effect.

Table 1.2: Regression with Different Internal Patent Definition

	(1)	(2)	(3)	(4)
Dependent Variable:				
Internal patent share, defined with a self-citation share larger than or equal to the following values				
Threshold of internal patent	30% Fuzzy	30%	50% Fuzzy	50%
Independent Variable:				
(Lag) Cash Flow/Asset × (1986-1992) Sector External Financial Dependence				
...	-0.0720***	-0.131***	-0.101***	-0.109***
	(0.0229)	(0.0464)	(0.0322)	(0.0421)
×Crisis	-0.0407**	-0.0574**	-0.0209	-0.0886*
	(0.0206)	(0.0289)	(0.0388)	(0.0519)
Firm and Sector Controls	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Firm FE	Correlated	Correlated	Correlated	Correlated
Instrument Variables	Y	Y	Y	Y
Observations	6,217	6,217	6,217	6,217

[a.] *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Sample period: 2002-2016. All variables (including controls) are winsorized at 1% and 99%. All regressions adopt the specification as Model (5) in Table 1.1. For details of controls and variable definitions, see the footnotes to Table 1.1.

[b.] Self-citation: when a patent cites its assignee's prior patents. Self-citation share for a patent = self-citations/total citations made (a.k.a. backward citations). I define internal patent using four thresholds: 30%, 50%, 30% (fuzzy) and 50% (fuzzy): all four assign a unit weight to a patent (i.e., count as one internal patent) when its self-citation share, x, passes the threshold, A, whereas the latter two also assign fractional weight, x/A , when $x < A$.

Table 1.3: Robustness Check of Internal Innovation and Financial Constraints Regression (1/2)

	(1)	(2)	(3)	(4)
Dependent Variable:				
Internal patent share, defined as patents with a self-citation share larger than or equal to 50%				
Independent Variable:				
(Lag) Cash Flow/Asset × …				
… (86-92) Sector External Financial Dependence	-0.103*	-0.117***		
	(0.0562)	(0.0451)		
… Crisis × (86-92) Sector External Financial Dependence	-0.412**	-0.0849**		
	(0.171)	(0.0424)		
… (86-92) EFD Dummy			-0.266***	
			(0.0833)	
… Crisis × (86-92) EFD Dummy			-1.140**	
			(0.502)	
… (86-01) Sector External Financial Dependence				-0.0559**
				(0.0234)
… Crisis × (86-01) Sector External Financial Dependence				-0.0294
				(0.0222)
Wald Test of Exogeneity (p-value)	0.0655	0.1014	0.163	0.0711
Observations	6,217	9,366	6,217	6,217
Number of Firms	1,365	1,708	1,365	1,365
Sample Period	2002-2016	1992-2016	2002-2016	2002-2016

[a.] *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. All variables (including controls) are winsorized at 1% and 99%. All regressions apply correlated random effect IV Tobit model as Model (5) in Table 1.1. Model (1) includes the full interaction; (2) expands the sample to 1992-2016; (3) uses external financial dependence (EFD) dummy and DID; (4) uses 1986-2001 EFD.

[b.] For details of controls and variable definitions, see the footnotes to Table 1.1.

Table 1.4: Robustness Check of Internal Innovation and Financial Constraints Regression (2/2)

	(5)	(6)	(7)	(8)
Dependent Variable:				
Internal patent share, defined as patents with a self-citation share larger than or equal to 50%				
Independent Variable:				
(Lag) Cash Flow/Asset × ...				
... (86-92) Sector External Financial Dependence	-0.148***			
	(0.0409)			
...		0.0807		
		(0.0611)		
... Crisis		-0.132***		
		(0.0477)		
... Dividend Dummy			-0.0939	
			(0.149)	
... Crisis × Dividend Dummy			-0.497*	
			(0.278)	
(Lag) Cash Flow/Net PP&E × ...				
... (86-92) Sector External Financial Dependence				-0.000445*
				(0.000226)
... Crisis × (86-92) Sector External Financial Dependence				-0.0129***
				(0.00328)
	0.0251	0.1122	0.023	0.135
Wald Test of Exogeneity (p-value)	6,217	6,217	6,217	6217
Observations	1,365	1,365	1,365	1,365
Number of Firms	2002-2016	2002-2016	2002-2016	2002-2016
Sample Period				

[a] *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. All variables (including controls) are winsorized at 1% and 99%. All regressions apply correlated random effect IV Tobit model as Model (5) in Table 1.1. Model (5) drops the interaction with crisis dummy; (6) drops the interaction with EFD; (7) uses a dividend dummy; (8) uses cash flow to net pp&e ratio. The detailed explanation is in the main text.

[b] For details of controls and variable definitions, see the footnotes to Table 1.1.

Similarly, α_1 and α_2 are reported.

I then change the calculation of EFD_s using 1986 to 2001 data instead of 1986 to 1992 as the main specification. α_1 is still significant at a 5% level.

In Column (5), I drop the Crisis_t from the interaction term in regression equation (1.1), essentially forcing the coefficient of $\text{EFD}_s \times \text{Cash Flow Ratio}_{i,t-1}$, i.e., α_1 to be time-invariant. As expected, the estimate is negative and statistically significant.

For (6), I drop the EFD_s in the interaction and only interact $\text{Cash Flow Ratio}_{i,t-1}$ with Crisis_t. In this way, I do not differentiate firms according to their external financial dependence. Instead, I analyze how the Great Recession, as a natural experiment, altered the relation between financing constraints and R&D for a given firm. The coefficient of $\text{Cash Flow Ratio}_{i,t-1}$ is insignificant, suggesting that on average, there is not strong evidence of financial friction's effect on R&D trajectory. This justifies controlling for external financing dependence in the main specification to identify their link (as mentioned in [AAB12]). The coefficient of $\text{Cash Flow Ratio}_{i,t-1} \times \text{Crisis}_t$ is still negative and significant at a 1% level, indicating that the financial condition worsened during the crisis.

For (7), I use the dividend payout pattern of a firm to determine when a firm is financially constrained. Following [BM06], I construct a dummy variable, $\text{Div}_{i,t}$, that equals zero when dividends are positive in both adjacent periods, and one otherwise. $\text{Div}_{i,t} = 1$ therefore signals an abrupt cessation in dividend payout, suggesting financial hardship. The regression equation is

$$\begin{aligned}
\text{Internal Patent Share}_{it} =& \alpha_0 + \alpha_1 \text{Div}_{i,t} \times \text{Cash Flow Ratio}_{i,t-1} \\
&+ \alpha_2 \text{Div}_{i,t} \times \text{Cash Flow Ratio}_{i,t-1} \times \text{Crisis}_t \\
&+ \alpha_3 \text{Cash Flow Ratio}_{i,t-1} \times \text{Crisis}_t + \alpha_4 \text{Cash Flow Ratio}_{i,t-1} \\
&+ \gamma' \text{Firm and Sector Controls}_{i,t-1} + u_i + v_t + \epsilon_{it}.
\end{aligned}$$

As shown in Table 1.3, α_2 is negative and significant. Financially stressed firms during the crisis indeed changed their R&D to be more internally oriented.

One advantage of (7) is that $\text{Div}_{i,t}$ is now firm and time specific, allowing firms within the same sector to be categorized differently in terms of financial constraint, and enabling firms to switch in and out of constraints. Yet a major drawback is that most innovative small firms do not pay dividends at all in the whole sampling period (so $\text{Div}_{i,t} = 0$ for all t), and many firms that stopped dividend payout (so $\text{Div}_{i,t} = 1$ for all t) are mature and less innovative. In addition, firms may reduce dividend payout rather than stop it completely. Due to these drawbacks, the main specification applies external financial dependence.

Lastly in (8), I change the definition of cash flow ratio. It is now calculated cash flow to net PP&E ratio. The conclusion stays unchanged.

1.6 Conclusions

This chapter investigates the empirical relationship between financial constraints and the type of firm innovations: internal (improve existing products) versus external (develop new product lines) innovation. The analysis, using matched firm-patent panel data of U.S. public firms from 1997 to 2016, reveals that firms become more inclined towards internal innovation when financially constrained, especially during the Great Recession. As external innovation has contributed more than internal innovation to U.S. economic growth, at least during the 1982-1997 period estimated by [AK15], and is an essential channel through which small firms grow, the alteration in innovation type can have profound implications on economic growth and firm dynamics. Hence, the findings imply that studies of the effects of financial frictions on firm innovation and growth should consider this new dimension in which innovation responds to financial constraints.

CHAPTER 2

The Impact of Financial Frictions on Innovation and the Size-Growth Relationship among Heterogeneous Firms

2.1 Introduction

The relationship between firm size and firm growth plays a central role in several classes of growth models and also in many models of firm dynamics. Most models assume a fixed relation: Some are based on Gibrat's law[1] that a firm's growth rate is independent of its size, while others feature a structure in which small firms grow faster than large firms[2]. Empirically, however, the firm size-growth relation is shown to be dependent on financial market conditions: Small firms grow faster[3] unless financial frictions are high, in which case we observe Gibrat's law[4][5]. This paper is among the first, to the best of my knowledge, to develop a unified framework that can account for both sets of observations, in which the relationship between firm growth and size depends critically on financial frictions—how difficult it is for firms to obtain outside financing.

The interaction between firm size-growth relation and financial frictions affects how firms innovate, which in turn influences economic growth. By considering both firm

[1]See [KK04],[LM08] and [AC15].

[2]See [AK15] and [AAA17]

[3] See [Sut97, Cav98, LM08, AK15].

[4] See [LSN10, LVS05, Cho10, FDA04, FM00, AA90]. For a literature review, see [NAA14, SKT06]; and Chapter 4 of [Coa09].

[5] The literature provides empirical evidence of the effect of financial frictions/development on firm growth, in both advanced and emerging economies. Some explicitly study Gibrat's law, such as [LVS05, Van05, AE06, AE10, LSN10]. Others examine the impact of finance on small firm growth more specifically, as in [KLC16, BDL08, CP02, BT02, BD06, BDM05]. [SV07] also offer a survey on how credit constraints limit the survival and growth rate of small firms.

dynamics and its effect on innovation, this framework allows me to analyze the aggregate implications of financial frictions with a much richer set of economic forces that that have been studied only in isolation. How small and large firms react differently to frictions? How would they adjust not only the amount but also the type of R&D they perform? How will firm size distribution evolve?

The model is an extension of [AK15] with financial constraints. Firms grow by performing two types of R&D: internal—improving existing products—and external—developing new products[6]. Subject to financial constraints, a firm's R&D expenditure is restricted by its internal funding: R&D costs cannot exceed a multiple (termed "pledgeability") of firm profits (so profits function as collateral).

In the model, financially constrained firms reduce their overall R&D expenditure and increase the share of internal innovation. The latter shift in R&D trajectory stems from the fact that external R&D is more elastic than internal R&D (due to, for instance, higher fixed costs). As a result, external R&D decreases proportionally more than internal R&D when constraints worsen.

The tightness of the financial constraint dictates the size-growth relation. On the one hand, when financial frictions are low, R&D is not (too) restricted by its cost. Assuming that a firm's ability in external R&D does not increase proportionally (linearly) with its number of product lines, small firms will thus be better at introducing new products (relative to their size) and grow faster, even if all firms are equally skilled in internal R&D[7]. On the other hand, R&D can be significantly constrained when financial frictions are high. This has two consequences. First, internal R&D share increases so small firms lose their comparative advantage; second, the overall level of R&D decreases, especially for small firms. Since external R&D is undirected, its cost depends on the aggregate level of productivity[8], so a less productive firm with lower profit will find R&D hard to afford.

[6]An example of the first kind is Apple's upgrade of iPhone each year, and of the second is Apple's introduction of a new electric vehicle business line (to compete with Tesla). This paper features creative destruction. Therefore, when a firm develops a new product, it captures market share from other firms.

[7] Intuitively, external R&D is not perfectly scalable because it is hard to use the knowledge/expertise embedded in existing product lines for new product development. For instance, compared to a start-up, Apple may be better at smartphone innovation, but no more capable at electric car engineering.

[8] For example, when Apple wants to develop electric vehicles (EVs), its R&D costs depend on the quality (and cost) of EVs in general, not iPhone's quality or costs.

Such firms are also likely to be smaller. With these two results combined, small firm growth rates are disproportionally reduced.

The model is consistent with several empirical observations made after the Great Recession. First, as described in Chapter 1, firms, especially small firms, reduce their R&D expenditure and switch toward internal innovation when financial constraints tighten. In addition, I document a drastic change in the size-growth pattern for U.S. innovative firms after 2007. Before 2007, there is a negative and significant relation between firm size and growth rate. However, this relation becomes insignificant after 2007. The result persists even after accounting for survival bias (Heckman), measurement error, and firm life cycle. This is consistent with the empirical literature that firm growth rate is affected by financial market conditions.

Before performing quantitative counterfactual analysis, I solve the equilibrium numerically and estimate model parameters by Simulated Method of Moments. The model solution is challenging. Because external R&D is undirected, firm value functions depend on the entire productivity distribution, which is an infinite-dimensional object characterized by a system of integro-differential delayed equations influenced concurrently by firm decisions. The solution method is inspired by reinforcement learning. In essence, I expand the state space for the value function to include firm-specific states, endogenous variables such as growth rate, to-be-estimated parameters, and discretized productivity distribution density: in total, 121 dimensions. The value function is then solved as a neural network by iteration over the Hamilton-Jacobi-Bellman equation.

I consider three adverse scenarios for the economy: a 5% drop in profit pledgeability[9], a 5% drop in aggregate demand (which reduces profits), and both. In the third case, the size-growth relation and labor productivity growth rate match well with their empirical counterparts in 2007–2016, suggesting a 22.8% drop in consumption-equivalent welfare if such counterfactual adverse drops are permanent.

Based on the third scenario with both drops, I analyze the effects of two policies. One is a 10% R&D subsidy to all firms, and the other is size-dependent. The second policy has an ex ante average subsidy rate of 10%, but a higher rate for smaller firms. The uniform

[9]Recall that R&D costs cannot exceed a multiple of profits. I estimate that the latter was reduced by 5% in 2007–2016 compared with 1997–2006.

subsidy improves the welfare to 82.3% of the baseline economy and the size-dependent subsidy to 88.5%. This suggests that offering more support to smaller firms, which are both more productive at R&D and susceptible to financial market conditions, can be beneficial.

2.2 Related Literature

This paper provides a unified theoretical framework in which the size-growth relation varies with financial frictions. The Schumpeterian growth literature predicts either proportional growth, such as [KK04] or a faster growth rate for small firms, as in [AK15] and [AAA17]. This paper offers a reconciliation of these two strands of models by introducing a constraint on R&D expenditures, which reacts to changes in demand or financial market conditions[10].

Related to the study on the aggregate impact of financial frictions, this paper contributes by incorporating a rich set of firm heterogeneity: R&D type, firm size, and firm productivity. Earlier papers consider a representative firm setting[11]. More recently, some authors stress the important role of firm heterogeneity in amplifying the impact of shocks. For example, [Gar16] considers how firms with different intangible capital to capital ratios react to asset price shocks in an economy with collateral constraints. However, his model does not feature endogenous growth. More closely related to my paper is [Sch17]. We both consider heterogeneous firms and imperfectly scalable R&D technology, but our models differ in two ways. First, he studies a one-time financial shock by which firms are cut out of financial markets and must rely on self-financing, while my definition of friction is more general: The pledgeability of cash flow is a continuous variable. Therefore, in addition to analyzing the aftermaths of a financial crisis, my model is useful for comparing countries with different financial market development. Second, I incorporate both internal and external R&D—another important dimension of heterogeneity—because the

[10]Empirically, this paper also provides evidence supporting the view that the validity of Gibrat's law is dependent on the economic environment (see footnotes 4, 5, and 3 for a brief review).

[11] [CG06] and [ACG16] study R&D in an expanding variety framework, and [Bar07] considers a quality ladder model. Some papers study the effect of shocks and/or frictions on human/physical capital accumulation and growth, including [AAB10] and [BKM19].

two types of R&D make distinct contributions to aggregate growth [AK15].

This paper is part of the nascent literature on solving complex continuous time heterogeneous agent models using machine learning techniques. [FHN18] use supervised learning tools to approximate the evolution of aggregate variables. [Dua18] applies reinforcement learning to solve value functions as neural networks. Building on [Dua18], I extend his method by considering a model in which endogenous distribution of agent types directly affects agents' decisions, and incorporate such a distribution into the domain of value function.

The rest of the paper is organized as follows. Section 2.3 presents three empirical findings and Section 2.4 presents the model. Several theoretical results are derived in Section 2.5. Section 2.6 quantifies the model and examines counterfactuals and R&D policies, and Section 2.7 concludes. Detailed proofs and the data description are included in the Appendix.

2.3 Firm Size-Growth Relation and the Great Recession

In this section, I document the empirical observations about the firm size-growth relation that motivate the model. The dataset is the same as in Chapter 1.

The last empirical fact pertains to the firm size-growth relation before and after the Great Recession in the U.S.. I find a negative relation between size and growth rate for all time periods except since 2007.

2.3.1 Baseline Regression

Following [AK15], I run the regression:

$$g_{f,t} = \alpha + \eta_{\text{Sector}\times\text{Year}} + \beta \log(\text{Employment}_{f,t}) + \epsilon_{f,t} \tag{2.1}$$

where $g_{f,t} = \frac{\text{Employment}_{t+1}}{\text{Employment}_t} - 1$ and $g_{f,t} = -1$ if firm f exits at time $t + 1$.

Table 2.1 contains the regression results. Column (1) and (2) refer to the estimates based on the 5-year or 10-year periods before 2007, while (3) and (4) on the post 2007

Table 2.1: Size-Growth Relation Before and After 2007

	Pre-Crisis		Post-Crisis	
	(1)	(2)	(3)	(4)
	2002-2006	1997-2006	2007-2011	2007-2016
	5-year	10-year	5-year	10-year
Log Employment	-0.0322**	-0.0336***	0.00142	-0.00668
	(0.0126)	(0.00749)	(0.00387)	(0.00457)
Constant	0.0137*	0.0188***	-0.0253***	0.0183***
	(0.00701)	(0.00650)	(0.00543)	(0.00588)
Observations	8,519	18,209	6,978	13,208
Sector-Year Controls	Y	Y	Y	Y

[a] *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Windsorized at 1% and 99%. Standard errors are clustered by firm.

counterparts. The estimates of β are negative and significant[12] prior to the Great Recession, indicating that small innovative firms grow faster than larger ones. However, such negative size-growth relation became statistically insignificant after 2007. In other words, we cannot reject Gibrat's law ($\beta = 0$) after the Great Recession.

The regression results echo the argument that size-growth relation is dependent on economic environment (see footnote 3), as firms found it harder to secure outside financing and suffered from diminished profits (inside financing) during the Great Recession. Both types of adverse development—tightened credits and dampened aggregate demand—will be reflected in the model as possible drivers of an increase in financial frictions.

2.3.2 Robustness Check

To test the robustness of the baseline results, I will control for firm survival bias, measurement errors and sample selections[13].

Table 2.2 presents the results. Column (1) is the baseline regression using 2002-2011 data with minor adjustment, given by

$$\begin{aligned} g_{f,t} = \alpha + \eta_{\text{Sector}\times\text{Year}} + \beta_1 \log(\text{Employment}_{f,t}) + \beta_2 \log(\text{Employment}_{f,t}) \times \text{Post 2007}_t \\ + \beta_3 \text{Post 2007}_t + \epsilon_{f,t} \end{aligned} \tag{2.2}$$

where indicator variable Post $2007_t = 1$ for $t \geq 2007$. β_1 measures the pre-2007 size-growth relation, $\beta_1 + \beta_2$ the post-2007 relation and β_2 the difference between the two. Their estimates are the same as those in (1) and (3) in Table 2.1. $\beta_2 > 0$ highlights the diminished size-growth relation after 2007.

I correct for the survival bias using the two-step Heckman selection model. In the baseline regression, when a firm exits the sample data, I set its employment growth rate $g_{f,t}$ to -1. However, this overestimates the actual growth rate employment before a firm exits, as -1 is the lower bound. Since small firms have higher exit rates, the

[12] [AK15] estimate β to be -0.0351 with standard error 0.0013 over the 1982-1997 period, which is very close to my estimates.

[13] I follow a robustness check similar to that in [Akc08]

Table 2.2: Robustness Check of Firm Size-Growth Relation

	(1)	(2)		(3)	(4)
Model	Baseline	Heckman	Select	IV	Age
Dependent variable: Employment growth rate					
Independent variables					
Log employment	-0.0322**	-0.0394***	0.0700***	-0.0292**	-0.0283**
	(0.0126)	(0.0102)	(0.0104)	(0.0132)	(0.0129)
Log employment × Post 2007	0.0336**	0.0384***	0.0254*	0.0316**	0.0339**
	(0.0132)	(0.00770)	(0.0135)	(0.0137)	(0.0132)
Log total long-term debt			0.00716**		
			(0.00354)		
Log debt due in 1 year			-0.0131***		
			(0.00394)		
Age					-0.00262***
					(0.000611)
Constant	-0.00388	0.0816	1.463***	-0.160*	0.0411***
	(0.00465)	(0.248)	(0.0242)	(0.0957)	(0.0114)
Observations	15,497	15,296 (1,162 censored)		15,497	15,497
Sector-Year FE	Y	Y		Y	Y
Instruments				Y	

[a.] *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Sample period: 2002-2011. All variables (including controls) are winsorized at 1% and 99%. Robust std. err. are clustered by firm in the parentheses. "Post 2007" is an indicator variable.

[b.] Model (1) is the baseline OLS with sector-year fixed effects. Model (2) uses the Heckman selection model, with the selection equation in the "select" column. Model (3) is 2SLS with lagged log employment as instruments. In Model (4), "age" is the year since a firm first appeared in Compustat (from 1970/1).

overestimation penalizes small firm growth rates, and thus β_1 is biased upward (closer to 0). To apply the Heckman selection model, I defined selected firms as remaining (surviving) firms. The selection equation is

$$\begin{aligned} 0 < & b_0 + b_1 \log(\text{Employment}_{f,t}) + b_2 \log(\text{Employment}_{f,t}) \times \text{Post 2007}_t + b_3 \text{Post 2007}_t \\ & + b_4 \log(\text{Total Long-Term Debt})_{f,t} + b_5 \log(\text{Debt Due in 1 Year})_{f,t} + u_{f,t}. \end{aligned} \tag{2.3}$$

Model (2) shows the Heckman regression results. The right sub-column presents the selection equation. The negative coefficient of log(Debt Due in 1 Year)$_{f,t}$ shows that, after controlling for the total indebtedness by log(Total Long-Term Debt)$_{f,t}$, firms with more short-term debts are associated with lower survival probability. Also, the positive b_1 estimate implies that large firms are more likely to survive.

From the left sub-column in (2), we see consistent results with the baseline regression. It shows a more negative β_1 as expected. In addition, a positive β_2 estimate indicates that small firms' growth rates were more negatively impacted after 2007.

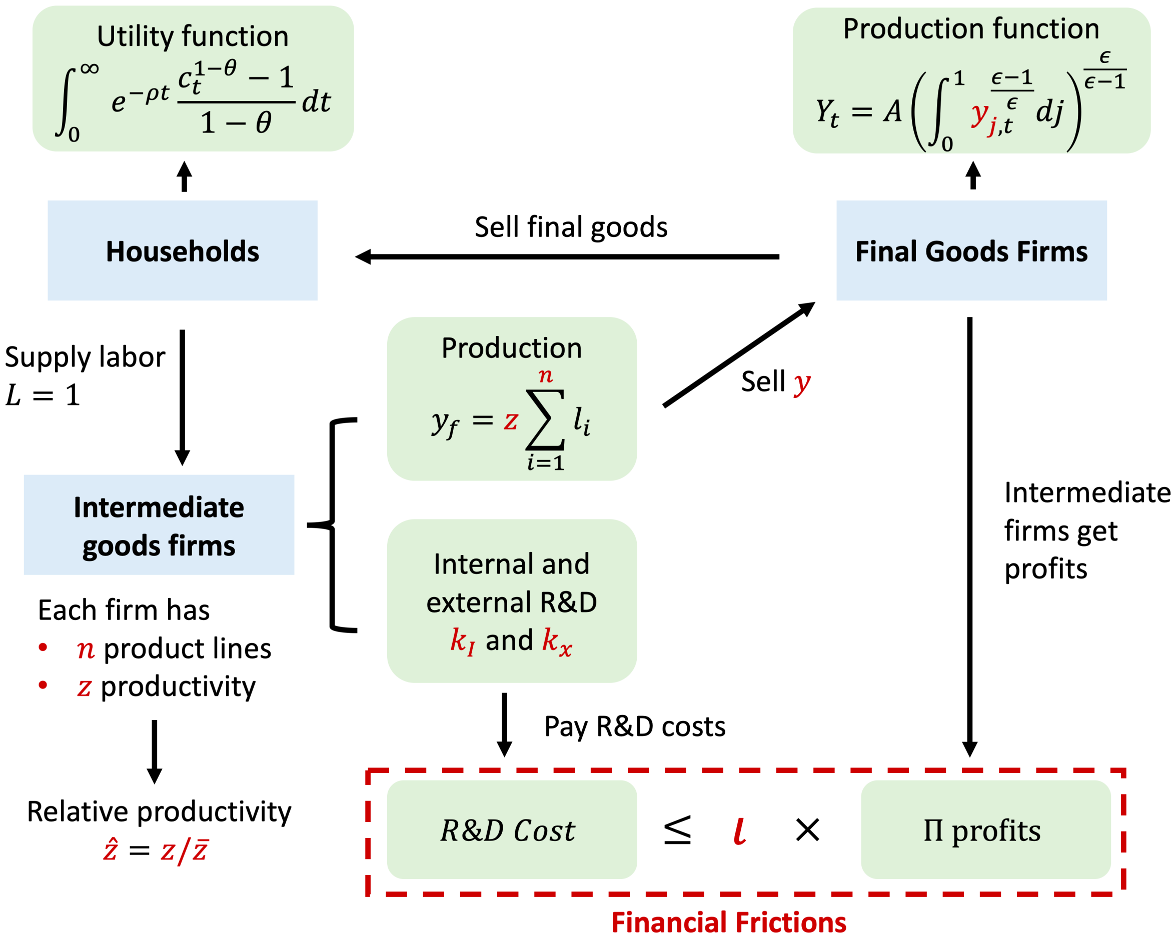
Column (3) accounts for possible measurement error in the proxy for firm size, i.e., firm employment. Since employment appears both in the regressors and denominator of the left-hand size variables, a measurement error can generate spurious negative association between size and growth rate and thus downward bias on β_2 (away from 0). I address this concern using lagged log(Employment) as instruments and apply a two-stage least square Model (2SLS). As expected, β_1 is closer to zero than Model (1). Nonetheless, the main conclusion remains unchanged[14].

Lastly, I address the sample selection problem in Column (4). Due to data limitation (Compustat), I include only public U.S. firms, so only 25% of the sample firms are considered typical "small" firms (less than 100 employees). This selection problem is solved by including age, a proxy of firm life-cycle, in the regression[15]. The negative

[14] The 2SLS model controlled for firm-clustered standard errors (SEs), which makes a standard overidentification test difficult to implement. I refit the model using heteroskedasticity-robust SEs (I obtain similar results as the cluster-robust SEs) and perform Wooldridge's score test of overidentifying restrictions. The p-value for the test is 0.2588, suggesting that the instruments are valid.

[15] Age is calculated as the number of years since a firm first appeared in the Compustat database.

Figure 2.1: Overview of the Model



coefficient of age indicates that firm growth rate decreases with age. The estimates of β_1 and β_2 remain relatively unchanged[16].

2.4 Model

In this section, I will build a model that can generate a different size-growth pattern depending on the tightness of financial frictions. The overview of the model is summarized in Figure 2.1.

[16]In addition, I run the regression with firm fixed effects to control for unobserved firm characteristics. The conclusions from the baseline model still hold.

2.4.1 Environment and Preferences

I consider a closed economy in a continuous time setting, admitting a representative household with CRRA utility function

$$U = \int_0^\infty e^{-\rho t} \frac{c(t)^{1-\theta} - 1}{1-\theta} dt \tag{2.4}$$

where $\rho > 0$ is the discount factor, θ is the constant relative risk aversion parameter and $c(t)$ is consumption at time t. The household is populated by a continuum of individuals with measure one. Each one is endowed with one unit of labor, $L = 1$, that is supplied inelastically at wage rate $w(t)$. The household also owns all the firms in the economy, which generates a risk-free flow rate of return $r(t)$. Therefore, the household maximizes (2.4) subject to the intertemporal budget constraint

$$c(t) + \dot{a}(t) \leq r(t)a(t) + w(t) \tag{2.5}$$

where $a(t)$ is the asset holdings of the household.

Individuals consume a final good $Y(t)$, which is also used for R&D (to be discussed later). The final good market is perfectly competitive. The final good is produced using a continuum of intermediate goods $j \in [0,1]$ according to the CES production technology:

$$Y(t) = A \left(\int_0^1 y_j(t)^{\frac{\epsilon-1}{\epsilon}} dj \right)^{\frac{\epsilon}{\epsilon-1}} \tag{2.6}$$

where $y_j(t)$ is the quantity of intermediate goods j, $\epsilon > 1$ is the elasticity of substitution between products, and A is the *aggregate demand parameter*—for reasons that will be made clear in Section 2.4.2. The final good $Y(t)$ is chosen as the numeraire.

2.4.2 Intermediate Goods Production and R&D

Measure M (endogenously determined) of firms produce intermediate goods. Each firm, indexed by $f \in M \in (0,1)$, owns a countable set of product lines $j \in \mathbf{J}_f \subset [0,1]$. A firm is characterized by the number of product lines n_f it operates (n_f is the cardinality of the set $\mathbf{J}_f$) and its labor productivity z_f. Each good $j \in \mathbf{J}_f$ is produced according to the linear technology

$$y_j = z_f l_j. \tag{2.7}$$

Compared with models such as [AK15] and [AAA17], I simplify the model by assuming the same productivity z_f for all $j \in \mathbf{J}_f$, instead of making the productivity good-specific (i.e., $z_j \neq z_i$ for $i, j \in \mathbf{J}_f$). With this simplification, I do not need to keep track of the multiset of productivity or the product portfolio of firm f. This will not affect the Schumpeterian (creative destruction) nature of the model, which will be explained in Section 2.4.2.2.

I define the aggregate labor productivity (or aggregate productivity for short) as

$$\bar{z} \equiv \left[\int_0^M \sum_{j\in\mathbf{J}_f} z_f^{\epsilon-1} d_f\right]^{\frac{1}{\epsilon-1}} = \left[\int_0^1 z_j^{\epsilon-1} d_j\right]^{\frac{1}{\epsilon-1}} \tag{2.8}$$

and the relative productivity by $\hat{z} \equiv z/\bar{z}$.

I omit the firm index f from now on when there is no confusion.

2.4.2.1 Profit Maximization

Each intermediate good producer is a monopolist[17]. The inverse demand function for each good j is

$$p_j = (Y/y_j)^{\frac{1}{\epsilon}} A^{\frac{\epsilon-1}{\epsilon}} = A\left(\frac{\bar{z}}{y_j}\right)^{1/\epsilon}. \tag{2.9}$$

Therefore, I interpret A as an *aggregate demand parameter.*

As shown in Appendix 4.3, in equilibrium, a firm characterized by (n, z) earns profits

$$\Pi(n, z) = \max_{\{y_j\},\{l_j\}_{j\in\mathbf{J}}} \sum_{j\in\mathbf{J}} [p_j y_j - w l_j] = n\bar{z}A\hat{z}^{\epsilon-1}/\epsilon. \tag{2.10}$$

[17] One can rationalize this assumption using a two-stage bidding game as [AK15].

Table 2.3: Internal versus External Innovation

	Internal R&D	External R&D
Example	iPhone X	Apple Car
n changes to	n	$n+1$
$\hat{z}$ changes to	$(1+\lambda)\hat{z}$	$\frac{n}{n+1}\hat{z}+\frac{1}{n+1}(1+\eta)z'$
Success rate	$\alpha_I k_I^{\beta_I}$	$n^{1+\gamma}\alpha_x k_x^{\beta_x}$
Research intensity	k_I	k_x
Elasticity	β_I	$\beta_x > \beta_I$
Return to scale	Constant	Decreasing as $\gamma < 0$
R&D cost	$\bar{z}n\hat{z}^{\epsilon-1}k_I$	$\mathbb{E}(n\bar{z}\hat{z}^{\epsilon-1}k_x) = n\bar{z}k_x$

[a] Innovation follows a Poisson process. z' is a random draw from existing $\hat{z}$ distribution $\phi(\hat{z})$. So z' is the original productivity of the product line being replaced (e.g., Nokia's productivity when iPhone replaced it).

2.4.2.2 Innovation

As is common in Schumpeterian models, the producer of each intermediate good j will be replaced by another firm (i.e., through creative destruction) unless it can keep innovating and successfully maintain its monopoly. There are two ways for a firm to innovate: internal R&D and external R&D. The former improves a firm's productivity and the latter expands its product lines by capturing markets from another incumbent. Table 2.3 summarizes the distinction between these two types of innovation.

Internal Innovation

Incumbents undertake internal innovation (a.k.a. R&D) to improve their labor productivity. Successful internal R&D arrives at the instantaneous Poisson rate

$$I = F_I(k_I) = \alpha_I k_I^{\beta_I} \tag{2.11}$$

where k_I is the internal research intensity, $0 < \beta_I < 1$ is the internal innovation elasticity, and $\alpha_I > 0$ is a scalar. Conditional on a successful internal innovation, firms' n and z update according to

$$n \to n \text{ and } z \to (1+\lambda)z$$

where $\lambda > 0$ is a multiplicative factor for productivity improvement.

External Innovation

Firms can also conduct external R&D to develop products that they do not currently produce[18]. External innovations are realized with the Poisson flow rate

$$X \equiv nx_n = nF_x(k_x, n) = n\alpha_x k_x^{\beta_x} n^{\gamma} \tag{2.12}$$

where k_x can be interpreted as external research intensity, $\alpha_x > 0, 0 < \beta_I < \beta_x < 1$ and $-1 \leq \gamma \leq 0$.

The condition $\beta_I < \beta_X$ implies that the success rate of external R&D is more elastic to R&D intensity (R&D inputs). One interpretation is that external R&D will succeed only when its research intensity is higher than some thresholds[19], making the innovation outcome very sensitive around these thresholds.

Another key assumption is $-1 \leq \gamma \leq 0$, following [AK15]. As $\gamma \geq -1$, the rate of external R&D is increasing in n—the number of product lines—because firms with more product lines are likely to have expertise in multiple fields, raising their odds of developing new products[20]. In addition, $\gamma \leq 0$ suggests that external R&D production

[18] In Schumpeterian models such as this one, the economy-wise variety of products is constant. Firms compete to obtain market from other incumbents.

[19] This is quite reasonable considering one needs to develop products in a completely new field/market. A firm will have to spend a certain amount in R&D, such as hiring new researchers, conducting market surveys, and making general administrative adjustments, before it can develop a new product line.

[20] As suggested by [KK04], n captures the human capital embedded in product lines.

function has a *diminishing return to* n (a.k.a. *imperfect scalablility*), because a firm's skill in *new* product development does not increase *linearly* with its expertise in its *current* product lines [21] (see also in footnote 7.).

Conditional on a successful external innovation,

$$n \to n + 1 \text{ and } z \to \frac{nz}{n+1} + \frac{z'(1+\eta)}{n+1}$$

where z' is the new product line's original producer's labor productivity (recall that a firm has to capture market from an incumbent) and $\eta > 0$.

It should be noted that the firm's new productivity is a weighted average between its original z and new $z'(1+\eta)$. A firm's *total labor productivity* nz always increases after external R&D[22], now becoming $nz + z'(1+\eta)$. As a result, the economy-wise aggregate productivity $\bar{z}$ is always increasing. However, the *average labor productivity* z for a firm can decrease if $z' \ll z$. This allows the model to capture the *dilution effect* of introducing a new product line, which is neglected by other Schumpeterian models[23].

Because external R&D is undirected across of all product lines $j \in [0, 1]$, a firm has equal probability to capture the market from any other incumbents. As a result, z' is a random draw from the current (endogenous) distribution of labor productivity. Furthermore, since each firm's product portfolio is of measure zero, firms will not innovate over their own product lines through external innovation.

2.4.2.3 R&D Cost Function

From Equation (2.10) and Section 2.4.2.2, it is clear that we can equivalently characterize a firm by n and its relative labor productivity $\hat{z}$.

The cost of internal innovation in units of final goods is $\bar{z}n\hat{z}^{\epsilon-1}k_I$. It is proportional to n and k_I, but convex in relative productivity $\hat{z}$ (strictly convex when $\epsilon > 2$).

---

[21] This is also supported by data. See [Coh10] for a literature review. I also provide some empirical evidence in 4.1.

[22] The incumbent who loses this product line has total labor productivity $(n' - 1)z'$, decreased from $n'z'$.

[23] For a firm with very high productivity, it is reasonable to consider the possibility of lowered average productivity as a result of new product development. For example, one can think of Amazon's Fire Phone, Google's Google Glass or Google Plus, and Apple's Newton. The cessation of these product developments all increased the corresponding firm's stock price.

To derive the cost of external innovation, we take expectation over firm specific productivity $\hat{z}$ and obtain $\bar{z}n\mathbb{E}(\hat{z}^{\epsilon-1})k_x = \bar{z}nk_x$, because external R&D is undirected (by construction, $\mathbb{E}(\hat{z}^{\epsilon-1}) = \int_0^1 \hat{z}_j^{\epsilon-1}dj = 1$). This implies that external R&D cost depends only on the *aggregate productivity*[24].

In total, a firm choosing innovation intensity k_I and k_x pays

$$R(k_I, k_x; n, z) \equiv \bar{z}n\hat{R}(k_I, k_x; \hat{z}) \equiv \bar{z}n[\hat{z}^{\epsilon-1}k_I + k_x] \tag{2.13}$$

units of final goods for R&D expenditure.

2.4.3 Financial Frictions

In this model, R&D expenditure cannot exceed a multiple of profits, or equivalently cash flow. This constraint is formally given by

$$R(k_I, k_x; n, z) \leq \iota\Pi(n, z) \tag{2.14}$$

where $\iota > 0$ is the *pledgeability* of profits. $\iota = \infty$ corresponds to a perfect capital market.

Empirical studies have stressed the importance of internal finance (e.g., cash flow or retailed earnings, both are related to profits), which has remained as the majority ($> 60\%$) source of funds for firms of all sizes [FHP06]. It is less affected by asymmetric information concerns common in external financing[25], and thus becomes essential for R&D—a type of intangible investment with high uncertainty[26].

Constraint (2.14) reflects the financial frictions (financial market imperfections) in the economy. Without any friction, firms can borrow indefinitely and thus afford any amount of R&D expenditure (i.e., $\iota = \infty$). When there is friction, a firm's debt capacity is limited. This limit is linked with firm profits in this model, because profits convey information of a firm's ability to generate cash flow and honor debt repayment[27].

[24] In standard Schumpeterian models as [KK04], there is only external R&D and its cost is dependent only on aggregate productivity. See also in footnote 8.

[25] See [MM84, Mye06] and many other related papers.

[26] See [BFP09, HL10, BP11, KN15].

[27] Also, the limit is only linked with profits, rather than a direct function of the sufficient statistics of

Constraint (2.14) is analogous to the collateral constraints in the literature[28], which states that total debt cannot exceed a multiple of capital (i.e., collateral). Firms do not possess capital in this model, so profit flows serve the purpose of collateral.

With (2.10) and (2.13), (2.14) can be simplified into $\bar{z}n[\hat{z}^{\epsilon-1}k_I + k_x] \leq \iota n\bar{z}A\hat{z}^{\epsilon-1}/\epsilon$. In other words,

$$k_I + k_x/\hat{z}^{\epsilon-1} \leq \iota A/\epsilon. \tag{2.15}$$

Immediate from (2.15), we can see that there are three ways for the financial constraint to tighten: (1) decrease in aggregate demand A; (2) decrease in profit pledgeability ι; and (3) decrease in relative productivity $\hat{z}$. The first two will affect every firm, while the third only small firms (in $\hat{z}$).

2.4.4 Discussion of Size-Growth Relation

There are two counteracting forces affecting firm size-growth relation: one from the innovation function (2.12) and the other from the financial constraint (2.15).

If financial friction is low (due to high A or ι), R&D expenditure will not be restricted by (2.15). Because of the diminishing return of n in (2.12), small firms (in n) grow faster and thus the size-growth relation is negative. This is the case analyzed in [AK15].

In contrast, if the constraint (2.15) is tight, then small firms' (in $\hat{z}$) innovation endeavors will be severely contained, while large firms' (in $\hat{z}$) will be relatively unrestricted. In addition, as firms lean toward internal innovation, which is perfectly scalable, small firms further lose their comparative advantage in innovation. As a result, small firms can no longer grow faster, and the economy features an independent (and even positive) size-growth pattern.

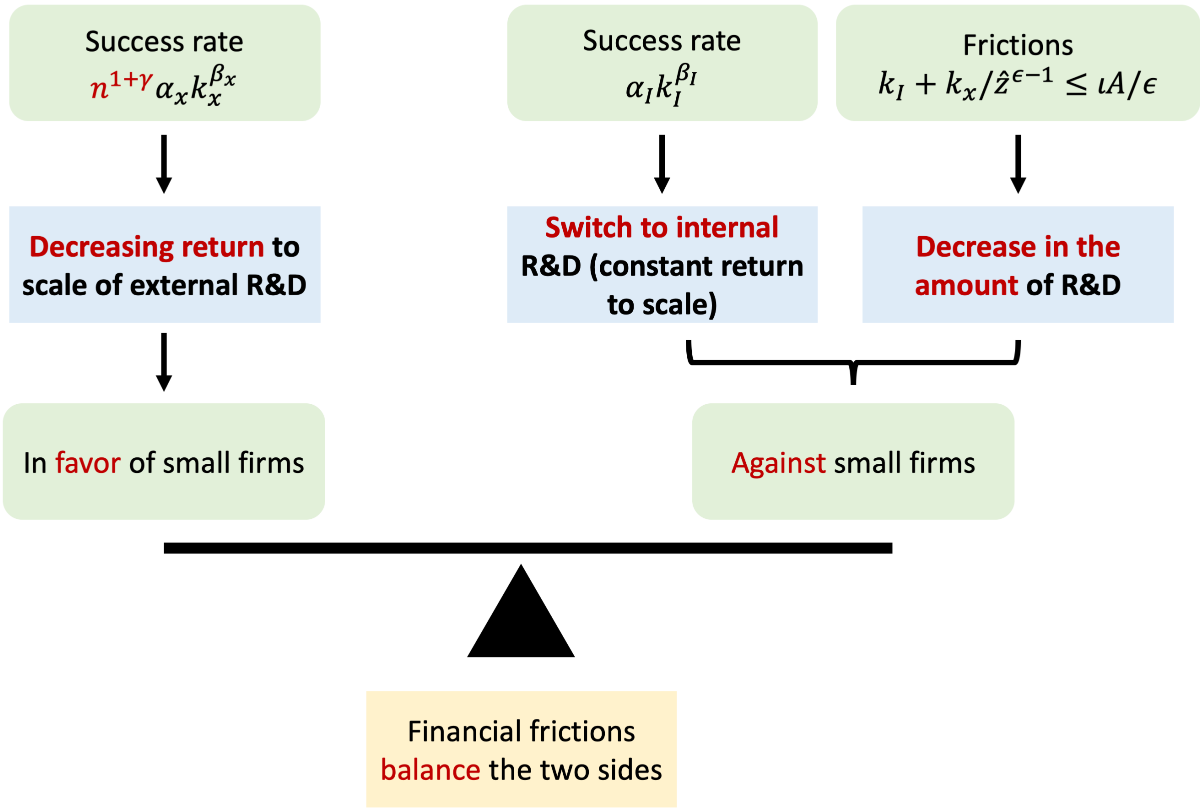
Which of (2.12) or (2.15) prevails then depends on the severity of financial frictions, as summarized by Figure 2.2. I will explore this result further in Section 2.5 and 2.6.

---

a firm: n and $\hat{z}$. One justification is that n and $\hat{z}$ are not observable (or verifiable) due to information frictions.

[28] [JQ12, Mol14, MX14, Gar16].

Figure 2.2: Illustration of the Model Mechanism



2.4.5 Value Functions

To summarize, a firm can be characterized by the pair $(n, \hat{z})$: the number of product lines and relative labor productivity. A firm's value will also depends on the aggregate productivity. Denote a firm's value function by $\mathbf{V}(n, \hat{z}, \bar{z})$. It follows the following lemma

Lemma 1 (Value Function). $\mathbf{V}(n,\hat{z},\bar{z}) = \bar{z}V(n,\hat{z})$ *and*

$$
\begin{aligned}
(r-g)V(n,\hat{z}) = \max_{k_I,k_x} \; & \underbrace{An\hat{z}^{\epsilon-1}/\epsilon}_{\textit{profit}} - \underbrace{n(\hat{z}^{\epsilon-1}k_I + k_x)}_{\textit{R\&D cost}} \\
& + \underbrace{F_I(k_I)[V(n,\hat{z}(1+\lambda)) - V(n,\hat{z})]}_{\textit{return from internal R\&D}} \\
& + \underbrace{nF_x(k_x,n)\left[\mathbb{E}_{\hat{z}'}V(n+1,\frac{n\hat{z}+\hat{z}'(1+\eta)}{n+1}) - V(n,\hat{z})\right]}_{\textit{return from external R\&D}} \\
& + \underbrace{n\tau[V(n-1,\hat{z}) - V(n,\hat{z})]}_{\textit{creative destruction}} \\
& - \underbrace{\frac{\partial V}{\partial \hat{z}}(n,\hat{z})g\hat{z}}_{\textit{change in }\hat{z}} + \underbrace{\varphi\left[\iota A\hat{z}^{\epsilon-1}/\epsilon - k_I\hat{z}^{\epsilon-1} - k_x\right]n}_{\textit{financial constraint}}
\end{aligned}
\tag{2.16}
$$

where τ is the equilibrium creative destruction rate, $g = \dot{\bar{z}}/\bar{z}$ and φ the Lagrangian multiplier of the financial constraint.

Proof. See Appendix 4.4 ■

Each incumbent firm maximizes its value by choosing R&D intensity k_I and k_x. The first line on the right hand side of (2.16) is the operating profit over currently held product lines net of R&D costs. The second line is the product of internal innovation Poisson arrival rate $F_I(k_I)$ and the change in firm value following internal innovation. The third line denotes the return from external innovation. The fourth line shows the change in firm value due to losing its product lines through creative destruction at Poisson rate τ. The first item in the final line represents the change in firm value due to the change in $\hat{z} \equiv z/\bar{z}$, as $\bar{z}$ grows at rate g. Lastly, we have the financial constraint faced by incumbents.

2.4.6 Entry and Exit

There is a unit measure of potential entrants. Each entrant has access to the external innovation technology similar to (2.12), different only up to a scalar. An entrant chooses an innovation flow rate $x_e > 0$ with cost (in terms of final goods) $\nu x_e^{\frac{1}{\beta_x}}\mathbb{E}[\hat{z}^{\epsilon-1}]\bar{z} = \nu x_e^{\frac{1}{\beta_x}}\bar{z}$, where $\nu > 0$ captures the entry costs. Upon a successful innovation, the entrant replaces an incumbent and starts producing intermediate goods. Denote the incumbent's (relative) labor productivity as $\hat{z}'$; the new entrant's productivity is $(1+\eta)\hat{z}'$. Since innovation is random, $\hat{z}'$ is drawn from the current (endogenous) distribution $\phi(\hat{z})$.

Denote $V(n, \hat{z})$ as the value of having n product lines with relative productivity $\hat{z}$, then entrants' maximization problem is

$$\max_{x_e} \bar{z} \left\{ x_e \mathbb{E}_{\hat{z}'} V(1, \hat{z}'(1+\eta)) - \nu x_e^{\frac{1}{\beta_x}}] \right\}.$$

Entry rate x_e then satisfies the free-entry condition

$$\mathbb{E}_{\hat{z}'} V(1, \hat{z}'(1+\eta))/\nu = x_e^{\frac{1}{\beta_x}-1}. \tag{2.17}$$

As in all Schumpeterian models, incumbents can lose some of their current product lines to other firms through competition (a.k.a. creative destruction). A firm that loses all product lines, i.e., $n = 0$, exits the economy.

2.4.7 Market Clearing and Stationary Distributions

I now close the model by specifying the market clearing conditions.

Final goods are used in consumption and innovation by incumbents and entrants. The market of final goods satisfies

$$Y = C + \bar{z} \int_{f \in M} n_f \hat{R}(k_I, k_x; \hat{z}_f) df + \bar{z}\nu x_e^{\frac{1}{\beta_x}}. \tag{2.18}$$

In terms of labor market,

$$\int_{f \in M} l_f df = \int_0^1 l_j dj = 1. \tag{2.19}$$

The equilibrium is also characterized by the stationary joint distribution of relative productivity $\hat{z}$ and firm size n. Denote the joint distribution by $H(n, \hat{z}) = Prob(\tilde{n} = n, q \leq \hat{z})$. Its density $h(n, \hat{z})$ (and also the measure of firms M) satisfies the Kolmogorov forward equation listed in Lemma 6 in Appendix 4.4. The marginal density of relative productivity $\hat{z}$ is then $\phi(\hat{z}) \equiv \sum_{n=1}^{\infty} h(n, \hat{z})$.

2.4.8 Aggregate Growth and Creative Destruction Rate

As shown in Appendix 4.3, the standard Euler equation states that

$$g = \frac{\dot{C}}{C} = \frac{\dot{Y}}{Y} = \frac{\dot{\bar{z}}}{\bar{z}} = \frac{r-\rho}{\theta}$$

We can also characterize the growth rate g and creative destruction rate.

Lemma 2 (Aggregate Growth and Creative Destruction)**.**

$$g = \frac{\overbrace{\tau\left[(1+\eta)^{\epsilon-1}-1\right]}^{\textit{external R\&D}} + \overbrace{\mathbb{E}_{\hat{z}}\left\{I(\hat{z})\left[(\hat{z}(1+\lambda))^{\epsilon-1}-\hat{z}^{\epsilon-1}\right]\right\}}^{\textit{internal R\&D}}}{\epsilon-1}. \tag{2.20}$$

where $I(\hat{z}) = M\sum_{n=1}^{\infty} F_I(k_I(n,\hat{z}))h(n,\hat{z})/\phi(\hat{z})$ is the internal innovation rate conditional on being type $\hat{z}$ firms, and M is the measure of incumbents.

The aggregate creative destruction rate is

$$\tau = \underbrace{M\sum_{n=1}^{\infty}\int_0^{\infty} nF_x(k_x(n,\hat{z}))h(n,\hat{z})d\hat{z}}_{\textit{agg. external R\&D rate}} + \underbrace{x_e}_{\textit{entry}}. \tag{2.21}$$

Proof. See Appendix 4.4. ■

Lemma 2 shows that growth comes from both internal and external innovation. External innovation is affected by the rate of creative destruction τ, which is jointly determined by both incumbents' and entrants' R&D efforts. Equation (2.20) also implies that the step sizes of R&D, η and τ, and the composition of internal/external R&D alter economic growth.

2.4.9 Stationary Equilibrium and Welfare

Finally, we can summarize the equilibrium of the economy.

Definition 1. *A stationary equilibrium consists of*

$$\{y_j, p_j, l_j, V(n,\hat{z}), k_x(n,\hat{z}), k_i(n,\hat{z}), x_e, M, h(n,\hat{z}), g, \tau, r, w\}$$

where

- *y_j, p_j and l_j maximize profit as in (2.10)*
- *$k_x(n, \hat{z})$ and $k_l(n, \hat{z})$ maximize $V(n, \hat{z})$. x_e solves entrants' problem as in (2.17)*
- *wage w is consistent with (2.18) and (2.19)*
- *the interest rate r satisfies the Euler equation $r = \rho + \theta g$*
- *The stationary distribution $h(n, \hat{z})$ and measure of firms M satisfy (4.4)*
- *g is given by (2.20)*
- *The creative destruction rate τ is given by (2.21)*

Normalize the initial aggregate productivity level to one; the welfare is given by the lifetime utility of the household

$$U(c_0, g) = \int_0^\infty e^{-\rho t} \frac{c(t)^{1-\theta} - 1}{1-\theta} dt = \frac{1}{1-\theta} \left[\frac{c_0^{1-\theta}}{\rho - (1-\theta)g} - \frac{1}{\rho} \right] \tag{2.22}$$

where

$$c_0 = \bar{z} \left[A - M \int_0^\infty \sum_{n=1}^\infty n(\hat{z}^{\epsilon-1} k_I(n, \hat{z}) + k_x(n, \hat{z})) dH(n, \hat{z}) - \nu x_e^{1/\beta_x} \right].$$

As in [AC15], I compare the consumption-equivalent change ξ along the balanced growth path for two economies with g^1, c_0^1 and g^2, c_0^2. ξ is defined as

$$U(\xi c_0^2, g^2) = U(c_0^1, g^1).$$

The equilibrium[29] is rather complex and has no analytic solution. In the next section, I will derive some theoretical results to clarify the intuition of the model.

[29]Note that the competitive equilibrium is inefficient, as in all Shumpeterian models, because firms do not fully internalize the positive externality of innovation. In addition, there is another source of inefficiency from the business stealing effect of external innovation [AAH14]. However, these inefficiencies are not the focus of this paper.

2.5 Theoretical Results

This section provides some theoretical results from the model and shows that the model is consistent with the three empirical facts documented in Section 2.3: (1) R&D growth rate decreases when financial constraints tighten, more severely for small firms (Section 1.3); (2) Constrained firms switch toward internal R&D (Section 1.4); and (3) The relation between size and growth rate will approach zero (i.e., independent relation) from negative when financial frictions increase (Section 2.3).

2.5.1 Firm Innovation and Financial Frictions

We have the following result regarding a firm's innovation intensity.

Lemma 3 (Firm Innovation Strategies)**.**

$$k_I^* = \left[\frac{\alpha_I \beta_I [V(n, \hat{z}(1+\lambda)) - V(n, \hat{z})]}{n\hat{z}^{\epsilon-1}(1+\varphi)}\right]^{\frac{1}{1-\beta_I}}$$
$$k_x^* = \left[\frac{\alpha_x \beta_x [\mathbb{E}_{\hat{z}'} V(n+1, \frac{n\hat{z}+\hat{z}'(1+\eta)}{n+1}) - V(n, \hat{z})]}{n^{-\gamma}(1+\varphi)}\right]^{\frac{1}{1-\beta_x}}$$

where φ is the shadow price of the constraint

$$\iota A\hat{z}^{\epsilon-1}/\epsilon \geq k_I \hat{z}^{\epsilon-1} + k_x.$$

Proof. See Appendix 4.4 ∎

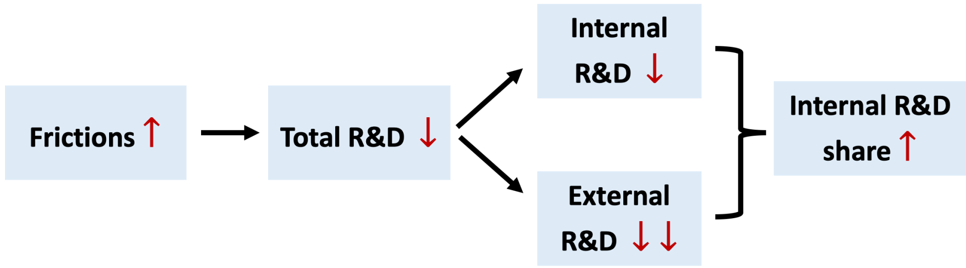
The $1+\varphi$ term in the denominator indicates that, everything else constant, the tighter the financial constrains (thus the higher the φ), the lower the innovation intensity k_I and k_x. This is related to the empirical fact analyzed in Section 1.3.

The next result pertains to Section 1.4.

Proposition 1 (Shift toward Internal Innovation)**.** *The ratio of internal to external innovation intensity $\frac{k_I}{k_x}$ is increasing in φ (decreasing in ι) when $\varphi > 0$.*

Proof. From Lemma 3, we have $\frac{k_I}{k_x} \propto (1+\varphi)^{\frac{1}{1-\beta_x} - \frac{1}{1-\beta_I}}$. Then the proposition follows naturally from $\beta_x > \beta_I$. See Appendix 4.4 for details. ∎

Figure 2.3: Shift to Internal Innovation



Recall that β_I and β_x are the elasticities of R&D production functions from equation (2.11) and (2.12). $\beta_x > \beta_I$ indicates that the success rate of developing new products is more sensitive to R&D inputs than internal innovation (the intuition is included in the texts after Equation (2.12)). Therefore, when total R&D expenditures drop, internal R&D will decrease more than external R&D. The ratio k_I/k_x will then decrease. Figure 2.3 illustrates the intuition.

The shift in the composition of R&D types has implications on the aggregate growth rate (see Lemma 2) and firm size-growth relation. The latter arises from the fact that internal R&D has constant return to n, so small firms are deprived of their advantages in innovation when they are forced to perform more internal innovation.

2.5.2 Firm Growth Rate and Financial Frictions

The section shows that the size-growth relation depends on financial frictions, relating to Section 2.3. From the equilibrium characterization in Appendix 4.3, a firm's size measured by employment is proportional to $n\hat{z}^{\epsilon-1} \equiv Q_f$. Accordingly, firm growth is equivalent to the growth of Q_f. We have the following definition:

Definition 2 (Firm Growth Rate). $g_f \equiv \mathbf{E}\left(\dot{Q}_f/Q_f\right)$ *is the average growth rate of a firm with* $n\hat{z}^{\epsilon-1} \equiv Q_f$. g_f *is characterized by Lemma 5 in Appendix 4.4.*

The relation between g_f and size Q_f (i.e., $\frac{\partial g_f}{\partial Q}$) depends on two opposing forces, as discussed in Section 2.4.4. On the one hand, the diminishing return of external R&D with respect to n (a.k.a. *imperfect scalablility*) works in favor of smaller firms (small in n). On the other, the financial constraint works against smaller firms (small in $\hat{z}$). Which force prevails depends on the severity of financial frictions (ι and A).

To clarify the intuition, consider myopic firms as in [AHS13], where firms consider only the current and immediate next period's profits, $\Pi(n, \hat{z})$. Also set $\epsilon = 2$ (linear profit function), $\gamma = -1$ (extreme inscalability) and normalize $A = 1$.

As shown in Appendix 4.5, when the financial constraint is not binding, i.e., $\varphi = 0$, $g_f = \frac{C_x}{n^{\frac{2-\beta_x}{1-\beta_x}} \hat{z}} + C_I - \tau$, where C_x and C_I are two constants[30]. Therefore $\frac{\partial g_f}{\partial Q} < 0$. This is the case analyzed in [AK15], in which smaller firms grow faster where only the first force, i.e., imperfect scalability, is at work.

However, when $\varphi > 0$, the sign of $\frac{\partial g_f}{\partial Q}$ is ambiguous (dependent on φ). When $\frac{\partial g_f}{\partial Q} = 0$, we will return to the case [KK04] where the growth rate follows Gibrat's law and is independent of size.

The above result is summarized in the following proposition:

Proposition 2 (Firm Growth Rate and Financial Frictions)**.** *Consider myopic firms with* $\epsilon = 2, \gamma = -1$.

- *When the financial constraint is not binding, i.e.,* $\varphi = 0$*,* $g_f = \frac{C_x}{n^{\frac{2-\beta_x}{1-\beta_x}} \hat{z}} + C_I - \tau$*, where* C_x *and* C_I *are two constants. The relation between size and growth rate is negative.*
$$\frac{\partial g_f}{\partial Q} < 0$$
- *When* $\varphi > 0$*, the size-growth relation is ambiguous.*

Proof. See Appendix 4.5. ■

2.6 Quantitative Analysis

This section estimates the parameters in the model using Simulated Method of Moments (SMM) and performs counterfactual and policy analysis.

[30] $C_x = \alpha_x(1+\eta)(\alpha_x\beta_x(1+\eta))^{\frac{\beta_x}{1-\beta_x}}$ and $C_I = \lambda\alpha_I(\alpha_I\beta_I\lambda)^{\frac{\beta_I}{1-\beta_I}}$.

2.6.1 Equilibrium Solution

Solving this model is challenging. The introduction of financial constraints (2.14) significantly increases the complexity of the model. For Schumpeterian models[31], a firm's value function is usually separable in terms of firms' different product lines. In other words, production (and innovation) on each product line is independent. However, the financial constraint is global, because it is determined by *firm-wide* profits and R&D expenditures.

In addition, methods such as finite-difference are typically computationally inefficient in this model. The main reason is that in external innovation, $\hat{z}$ is not contained to movement around its neighborhood[32]: it spans the whole distribution support according to $\hat{z} \to \frac{n\hat{z}}{n+1} + \frac{\hat{z}'(1+\eta)}{n+1}$.

The value function $V(n, \hat{z})$ (2.16) directly depends on the distribution of relative productivity $\phi(\hat{z})$, instead of a single-dimensional equilibrium variable (e.g., prices). Furthermore, the joint distribution of number of product lines and relative productivity $h(n, \hat{z})$ is determined by a complex system of delay integral-differential equations (see Lemma 6 in Appendix 4.4), which, in turn, are affected by firm value functions (firm choices).

To calculate the joint distribution, I discretize the state-space of n and $\hat{z}$ into a grid[33]. The discretized version of density $h(n, \hat{z})$ is the stationary distribution of a discrete Markov chain, of which the transition matrix depends on the firm's innovation choices: Firms move up and down along n or $\hat{z}$ dimension within the grid with reflective boundaries. Given the joint distribution, $\phi(\hat{z})$ is represented by a discretized probability distribution (I use 108 grid points for $\hat{z}$). The details are in Appendix 4.6.1.

The value function is approximated by a 3-layer neural network $V(n, \hat{z}, g, \tau, \phi(\hat{z}), \Omega)$. Ω are the 9 to-be-estimated parameters (explained later). The total input dimension is 121. The calculation of value functions is similar to [Dua18]. It explicitly requires the convergence of the Hamiltonian-Jacobian-Bellman (HJB) equation. Note the solution of $V(n, \hat{z}, g, \tau, \phi(\hat{z}), \Omega)$ is a function, so we can easily compute the value given any combination of its inputs without performing the value function iteration again.

[31]Such as [LM08, AHS13] and [AK15].

[32]Technically, the HJB equation will not produce a sparse matrix. See [AAM17]

[33]As in [AHS13], I impose upper bounds on n and $\hat{z}$ (10 and 6) in the numerical solution.

Table 2.4: Calibrated Parameters

	Description	Value	Notes
ρ	Discount rate	0.02	Annual discount factor 0.97
θ	Inverse of the inter-temporal elasticity of substitution	2.00	[AAA17]
ϵ	Elasticity of substitution between products	2.90	[AAA17]
A	Production function scalar	1.00	Normalized

I employ the Python packages TensorFlow and Keras to train the neural network (reinforcement learning). The details are given in Appendix 4.6.2.

The overall algorithm is summarized below.

Algorithm Overview

1. Calculate the neural network approximiation of value function $V(n, \hat{z}, g, \tau, \phi(\hat{z}), \Omega)$ and thus innovation rates $I(n, \hat{z}) = F_I(k_I(n, \hat{z}))$ and $X(n, \hat{z}) = nx(n, \hat{z}) = nF_x(k_x(n, \hat{z}), n)$.

2. Find the equilibrium for any given parameters Ω.

   (a) Guess equilibrium g and τ

   - Guess relative productivity distribution $\phi(\hat{z})$
   - Calculate innovation rates using the value functions
   - Calculate the stationary joint distribution $h(n, \hat{z}; g, \tau, \Omega)$
   - Update $\phi(\hat{z})$
   - Repeat until $\phi(\hat{z})$ converges.

   (b) Update g' and τ' according to (2.20) and (2.21).

   (c) Solve the fixed points of $g' = F_g(g, \tau; \Omega)$ and $\tau' = F_\tau(g, \tau; \Omega)$, where F_g and F_τ denote the functions that return the updated value g' and τ'. This step is solved using the Python non-linear function solver.

Table 2.5: Estimated Parameters

Parameter	Description	Estimated value
β_I	Internal innovation elasticity	0.408
β_x	External innovation elasticity	0.471
α_x	External innovation scalar	0.501
α_I	Internal innovation scalar	6.483
η	External innovation step size	0.099
λ	Internal innovation step size	0.018
γ	External innovation scalability	-0.463
ν	Entry cost scalar	2.482
ι	Profit pledgeability scalar	5.76e-4

2.6.2 Estimation

This section presents the estimation procedure of model parameters. Four parameters listed in Table 2.4 are externally calibrated to values common in the literature. The remaining nine parameters in Table 2.5, or Ω, are estimated by matching data moments.

I use Simulated Method of Moments (SMM) to estimate the parameters. Define $\Gamma(\mathbf{Y})$ and $\Gamma(\Omega)$ to be the data ($\mathbf{Y}$) moments and simulation moments. The estimator minimizes

$$\hat{\Omega} = \underset{\Omega}{\text{argmin}}[\Gamma(\mathbf{Y}) - \Gamma(\Omega)]^T \mathbf{W}[\Gamma(\mathbf{Y}) - \Gamma(\Omega)]$$

where $\mathbf{W}$ is a diagonal weighting matrix with $\mathbf{W}_{ii} = 1/\Gamma_i(\mathbf{Y})^2$ and $\mathbf{W}_{ij} = 0$ for $i \neq j$. The aggregate growth rate's weight is increased by a factor of three [AHS13]. The simulated moments $\Gamma(\Omega)$ are calculated by Monte Carlo with 5,000 firms.

All parameters are identified jointly. The next subsection provides a heuristic discussion on identification.

2.6.2.1 Empirical Moments and Identification

In the model, size is jointly determined by n and $\hat{z}$. Revenue, profit and employment are linear in $Q_f = n\hat{z}^{\epsilon-1}$. Since total employment $l_f = Q_f$ in equilibrium, I will use employment as the size measure.

Growth Regression Coefficient Γ_1

The empirical regression is given by Table 2.1. In simulation, it is the regression of firm growth rate $g_f(Q_f)$ on $\log(Q_f) \equiv \log(n\hat{z}^{\epsilon-1})$. As mentioned in the theoretical section 2.5, Γ_1 is particularly sensitive to γ and ι.

Firm Entry Rate Γ_2

Due to data limitation (no Census data), I will use the annual average firm death rate from Business Dynamics Statistics[34]. Entry rate is estimated as the annual average of employment share of age 0 & 1 firms. The model equivalent is $x_e\mathbb{E}(\hat{z}^{1-\epsilon})$. Entry rate will discipline ν, η, ι, α_x and β_x.

Aggregate Labor Productivity Growth Rate Γ_3

The aggregate labor productivity is calculated as real total value-added per worker. Value-added is the difference between net sales and cost of goods sold. Growth rates are computed as geometric averages. It is particularly affected by τ and innovation step sizes η and λ, as shown in equation (2.20).

Firm Growth Rate by Size Γ_4, Γ_5

They are defined as the averaged annual growth rates of total employment for firms larger or smaller than median size in the sample. They are informative in determining γ, τ, η and λ and innovation production function. The expression of $g_f(Q_f)$ is given by 4.3 in Lemma 5.

Firm Relative Productivity Growth Rate by Size Γ_6, Γ_7

The model equivalent is $\mathbb{E}[I\lambda + nx\frac{(1+\eta)\hat{z}'/\hat{z}-1}{n+1}] - g$, where I and x are (firm-specific) Poisson arrival rates of internal and external R&D.

Internal Innovation Ratio by Size Γ_8

I define Γ_8 as the ratio of internal innovation share between large and small firms, where internal innovation share is proxy by the share of internal patents using a fuzzy 30% cutoff. It corresponds to $\mathbb{E}(\frac{F_I}{nF_x+F_I})$ in the model.

[34] `https://www.census.gov/ces/dataproducts/bds/data_firm.html`. I exclude sectors 00-09 (AGR Agriculture, Forestry, and Fishing) and 60 (FIRE Finance, Insurance, and Real Estate). The time series plot is included in Appendix 4.2

Table 2.6: Model and Data Moments

Moments	Data	Model
Growth regression coefficient	-0.034	-0.040
Entry employment share	0.055	0.049
Aggregate labor productivity growth rate	0.030	0.027
Small firm employment growth	0.076	0.087
Large firm employment growth	0.063	0.045
Small firm productivity growth	0.069	0.091
Large firm productivity growth	0.009	0.006
Small/large firm internal patent share ratio	0.542	0.671
Small/large firm R&D intensity ratio	2.474	2.968

Innovation Intensity by Size Γ_9

Innovation intensity is measured as R&D expenditure to net sales ratio, or $\epsilon(k_I + k_x\hat{z}^{1-\epsilon})/A$. Γ_9 is informative on innovation production function parameters, ι and γ

2.6.2.2 Model Fit

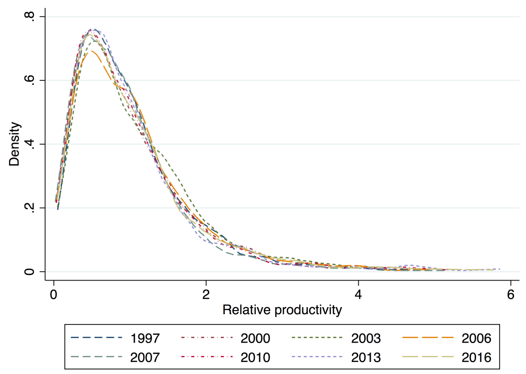
The estimation results are shown in Table 2.6. The simulated moments match the empirical counterparts relatively well.

Besides targeted moments, I also compare the untargeted moments, namely relative productivity distribution and firm size distribution.

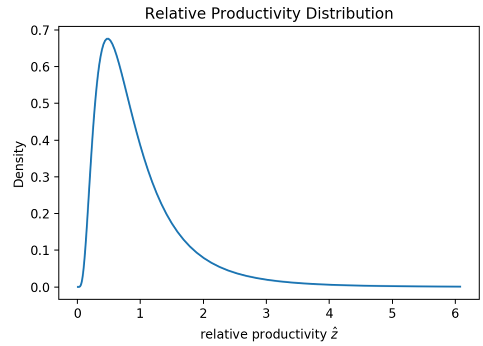
The model produces relative productivity[35] distribution similar to the empirical one, shown in Figure 2.4. On the left panel, I plot the distribution for each three years throughout the sample periods. It shows a stable distribution over time, even considering that the sample spans both the early 2000s Dot-Com bubble and the Great Recession. The model, however, cannot fully match the left tail of the empirical distribution. One explanation is that the Compustat dataset does not contain enough small firms. Also, the model does not feature fixed operation cost or exogenous exits, so the survival rates

[35] Relative productivity $\hat{z}_i = z_i/\bar{z}$. $\bar{z}$ = Total real value added/Total employment. Same for z_i.

Figure 2.4: Relative Productivity Distribution

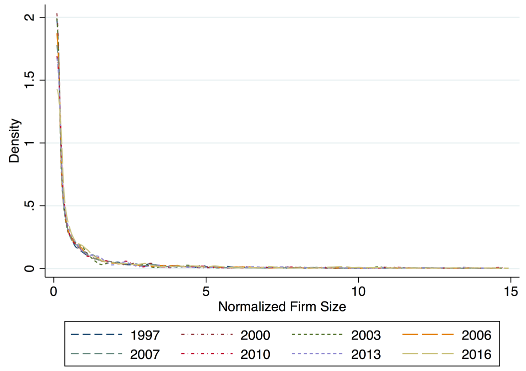


(a) Data

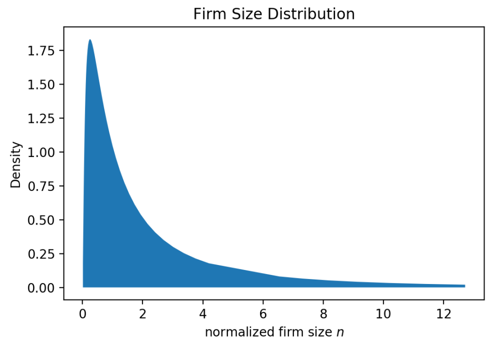


(b) Model

Figure 2.5: Firm Size Distribution (Normalized to Unit Mean)



(a) Data



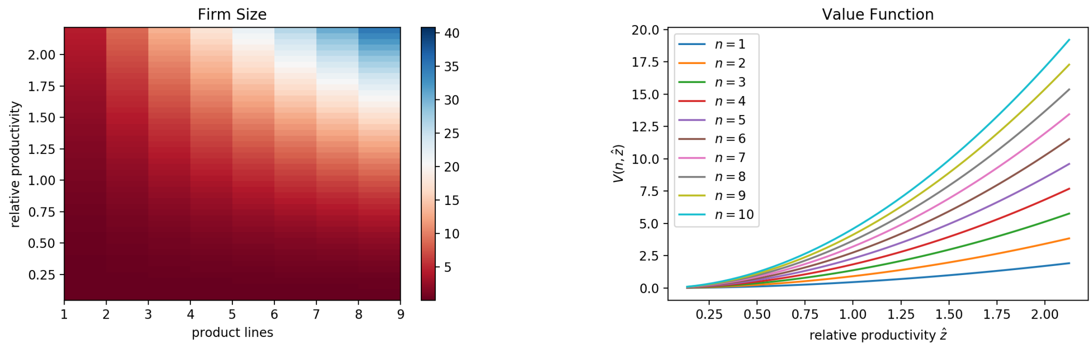
(b) Model

(and growth rates) of small firms can be higher from the simulation[36].

The simulated firm size distribution also resembles its empirical counterpart as in Figure 2.5. In both the data and model, firm size is defined as total employment, or $n\hat{z}^{\epsilon-1}$. To make the two comparable, I normalize both distributions to have unit mean. The empirical distribution takes the shape of a Pareto distribution, consistent with the literature (e.g., [KK04]). The simulated one has less mass concentrated on the left tail, which is inherited from the relative productivity distribution in Figure 2.5(b).

[36]/This is also apparent in Table 2.6. Compared to the empirical moments, the model generates higher (lower) growth rates (in both employment and productivity) for small (large) firms. The simulated growth coefficient, −0.040, is also more negative.

Figure 2.6: Firm Size and Value Function



(a) Firm Size (b) Value Function

2.6.2.3 Characterization of the Economy

In this section, I will discuss some properties from the simulated economy.

Value Function

For illustration, I plot the relation between size and $(n, \hat{z})$ in Figure 2.6, which is determined by $Q = n\hat{z}^{\epsilon-1}$.

A firm's value function, $V(n, \hat{z})$ in Equation (2.16), is plotted in Figure 2.6. $V(n, \hat{z})$ is increasing in both the number of product lines n and relative productivity $\hat{z}$.

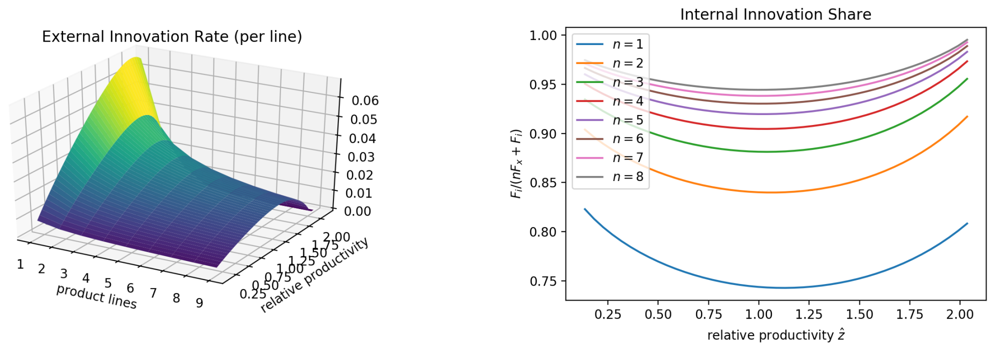
Innovation Rate

Figure 2.7 depicts the innovation policy function for different n and $\hat{z}$. External innovation rate per product line[37], $F_x(n, \hat{z}) = \alpha_x k_x^{\beta_x} n^{\gamma}$, is shown in the left panel. Across the n-axis, F_x decreases as the number of product lines increases. This reflects the comparative advantage in external innovation for firms with fewer product lines, captured by the n^{γ} term in the model.

Along the $\hat{z}$-axis, we see an inverted-U shape of F_x. The upwar-sloping fraction results from the financial frictions that R&D expenses cannot exceed a multiple of profits, $\iota A\hat{z}^{\epsilon-1}/\epsilon \geq k_I\hat{z}^{\epsilon-1} + k_x$. Therefore, given n, higher $\hat{z}$ will relax the constraint, enabling

[37] The external innovation intensity $k_x(n, \hat{z})$ shows a similar pattern.

Figure 2.7: Innovation Rate



(a) External Innovation Rate (b) Internal Innovation Share

more R&D and a higher innovation rate. This is a countervailing force against the inscalability of external innovation. It works in favor of firms large in terms of $\hat{z}$.

The downward-sloping part comes from the fact that after a successful external innovation, a firm's productivity $\hat{z}$ updates according to $\hat{z} \to \frac{n\hat{z}}{n+1} + \frac{\hat{z}'(1+\eta)}{n+1}$. A firm with very high $\hat{z}$ is unlikely to obtain a higher $\hat{z}'$, and thus the benefit of external innovation diminishes. This is the *dilution effect* analyzed in Section 2.4.2.2.

The internal innovation share, $\frac{F_I(n,\hat{z})}{F_I(n,\hat{z})+nF_x(n,\hat{z})}$, is shown on the right panel of Figure 2.7. Consistent with observations from the external innovation rate, smaller firms in terms of n have higher internal innovation share, similar to [AK15]. We also see a U-shaped relation between $\hat{z}$ and internal innovation share.

2.6.3 Counterfactual and Policy Analysis

In this section, I will perform several counterfactual analyses and policy experiments to quantify the implications of financial frictions, and to suggest how to design appropriate policies.

2.6.3.1 Quantifying Financial Constraints

The Great Recession is a period with sizable change in financial markets and aggregate economy. In Section 2.3, I have shown evidence that firm innovation and growth rate were affected by the crisis. I will now first quantify the change in financial constraints after the Great Recession, and then evaluate its implications.

From (2.15), constraints can become more severe for all firms when A or ι decreases. I refer to the changes to ι and A as profit pledgeability and aggregate demand shocks[38].

Frictionless Counterfactual

As a benchmark, I consider the case where the financial constraint never binds, i.e., $\iota = \infty$. The result is shown in Table 2.8. The productivity growth rate g is increased to 0.035, and the negative relation between firm size and growth rate (β) is more significant. The total welfare (in consumption-variation terms, see Equation (2.22)) increases slightly by 10.2% compared to the baseline case[39].

Decrease in Pledgeability ι

In the model, the financial friction is given by

$$R_t(k_I, k_x; n, z) \leq \iota \Pi_t(n, z).$$

In other words, $\frac{R_t(k_I,k_x;n,z)}{\Pi_t(n,z)} \leq \iota$. Motivated by the model, I estimate the ι for each sample period (1997-2006, 2007-2016) by

$$\bar{\iota} = \frac{1}{T} \sum_t \frac{1}{N_t} \sum_i \text{R\&D expenses}_{it} / \text{Cash Flow}_{i,t-1}. \tag{2.23}$$

Cash $\text{Flow}_{i,t-1}$ is defined the same as it is in the empirical regression (1.1). The estimated ι are listed in Table 2.7. So in the counterfactual analysis, I consider the case where ι is decreased by 5%.

Table 2.8 shows the equilibrium outcome after a 5% decrease in ι. The aggregate labor productivity growth rate drops by 0.6 percentage points compared to the baseline scenario. The relation between growth rates and firm sizes also weakens to -0.020.

[38]Even though I use the term "shocks," I do not perform impulse response analysis. Instead, I compare stationary equilibria (balanced growth paths) after permanent change in parameter values.

[39]Note that higher growth rates from more innovation efforts also entail higher R&D costs and thus less to consume.

Table 2.7: Estimation of Change in ι_t

	1997-2006	2007-2016	% Change
Average	0.325	0.311	-4.6%
SD	0.037	0.018	

[a.] R&D expenses$_{it}$/Cash Flow$_{i,t-1}$ are winsorized at 5% and 95% by year. Cash Flow$_{i,t-1}$ is calculated as income before extraordinary items plus depreciation and amortization.

Welfare also suffers a 8.6% decrease. However, both the growth rate g and coefficient β are still higher than those from the 2007-2016 data.

Decrease in aggregate demand A

I now consider a drop in A by the same magnitude 5%[40]. The consequence now is more severe than a 5% ι drop. The growth rate is now 0.9 percentage points lower and welfare 23.0% less. Also, the growth rate difference between large and small firms, measured by β, is significantly smaller at -0.007. One possible explanation is that A affects not only the financing constraint, but also the profit flow $\pi = n\bar{z}A\hat{z}^{\epsilon-1}/\epsilon$ and thus firm values. Reduced firm values depress both innovation, internal and external, and firm entry. Furthermore, smaller A also directly reduces final output from Equation (2.6). This result signifies the importance of boosting aggregate demand in recessions in addition to restoring a well-functioning financial system.

Decrease in both ι and A

More likely than not, both the financial market and aggregate demand worsened during the Great Recession. When both ι and A are reduced to 95% of their baseline levels, the aggregate productivity growth g is 0.010, very close to the actual 0.009 measured in the data. Similarly, $beta$ is -0.006, within the 95% confidence interval, $(-0.014, -0.004)$, of

[40]To put this into perspective, during the Great Recession, employment dropped by 6.7%, output by 7.2% and consumption by 5.4% in the United States from 2007 Q4 to 2009 Q3.

Table 2.8: Counterfactual Analysis

	iota	A	Beta	g	Welfare
Baseline	$\bar{\iota}$	1	-0.040	0.027	100
Frictionless	∞	1	-0.060	0.035	110.2
Decrease in profit pledgeability	$0.95\bar{\iota}$	1	-0.020	0.021	91.4
Decrease in aggregate demand	$\bar{\iota}$	0.95	-0.007	0.018	87.0
Decrease in both	$0.95\bar{\iota}$	0.95	-0.006	0.010	77.2
1997-2006 Data			-0.034	0.030	
2007-2016 Data			-0.009	0.009	

[a.] The baseline model is estimated using 1997-2006 data. In the frictionless case, $\iota = \infty$.

[b.] Beta refers to the regression coefficient of firm log employment on the firm employment growth rate. g is the aggregate labor productivity growth rate. The baseline welfare is normalized to 100.

that estimated from the data. The welfare loss is approximately 22.8% compared to the baseline economy.

2.6.4 R&D Subsidy

Given the 5% decrease in both ι and A, I consider two types of R&D subsidies aiming to relax the financing constraint. Denote the (firm-specific) subsidy rate by s_f. The financing constraint with subsidy is now

$$\frac{\iota A \hat{z}^{\epsilon-1}}{(1-s_f)\epsilon} \geq k_I \hat{z}^{\epsilon-1} + k_x. \tag{2.24}$$

The contemporaneous after-R&D profit flow is $A\hat{z}^{\epsilon-1}/\epsilon - n(1-s_f)(k_I\hat{z}^{\epsilon-1}+k_x)$. The subsidy is financed by a lump-sum tax T to the shareholders , i.e., the representative household where $T = \bar{z}\int_f s_f n_f \hat{R}(k_I, k_x; \hat{z}_f)df$.

Uniform Subsidy

As both ι and A decrease by 5%, I consider a simple policy that every firm receives a 10% subsidy to their R&D. Table 2.9 presents the results. Both the welfare and aggregate growth rate g are now much higher than before the subsidy. The size-growth relation also favors smaller firms more with $\beta = -0.012$, compared to -0.006 without subsidies. Intuitively, this is because now the financing constraint is essentially restored to the baseline case[41]. Nevertheless, all metrics are still worse than the baseline economy, since deteriorated aggregate demand will negatively affect firm values and welfare.

Size-dependent Subsidy

As advocated in the literature as well as implemented in practice (see for example [AAA17]), smaller firms should receive higher subsidy, as they are more susceptible to demand fluctuation and financial market imperfections. Here I consider a size-dependent policy where the average subsidy is still 10%. In particular, let

$$s_f = s(Q_f) = 1 - 1/(a + b\frac{Q_{min}}{Q_f}) \tag{2.25}$$

where $Q_f = n\hat{z}^{\epsilon-1}$ and $Q_{min} = \min_{f\in[0,M]} Q_f$. Note that $\frac{1}{1-s(Q_f)} = a + b\frac{Q_{min}}{Q_f}$. Also,

$$\sum_{n=1}^{\infty}\int_0^{\infty} s(Q)dH(n,\hat{z}) = 10\%$$
$$s(Q_{max}) = 0.$$

Although in the theoretical model, $Q_{max} = \infty$, in the numerical solution, I set $Q_{max} = n_{max}\hat{z}_{max}^{\epsilon-1}$. Also, I set $H(n,\hat{z})$ to be the distribution of the no-subsidy economy, so a more appropriate interpretation is that the ex-ante subsidy is 10%. a and b are two parameters, calculated to be 1.099980 and 35.132840 respectively.

As shown in Table 2.9, there is a modest improvement compared to the uniform subsidy case, albeit still less than the baseline scenario for the same reason discussed before. Consistent with the literature, a size-dependent subsidy is more effective than a uniform

[41] To see this, note that $0.95 * 0.95/0.9 \approx 1.0028$

Table 2.9: Policy Analysis

	Beta	g	Welfare
Uniform subsidy	-0.012	0.015	82.3
Size-dependent subsidy	-0.019	0.019	88.5

a. In both cases, ι and A decrease by 5% compared to the baseline case in Table 2.8.

one, considering smaller firms are associated with higher innovation capacity [AK15]. The improvement is less pronounced than in other estimates. Some studies further differentiate innovation efficiency among firms within the same size group (e.g., compare large-old, large-young, small-old and small-young firms). Other models introduce taxation on large but inefficient incumbents [AC15, AAA17]. Still other papers relate subsidy directly with firm productivity [Akc08]. In this paper, a large firm is not necessarily more productive, as firm size Q is a function of both the number of product lines n and productivity $\hat{z}$. Interestingly, even with this crude measure of size, size-dependent R&D policy still yields better results.

2.7 Conclusions

In this paper, I build a Schumpeterian growth model with financial frictions. Firms are financially constrained so that their R&D spending cannot exceed a multiple of profits. Depending on the severity of the constraint, size-growth relation can be negative or independent. Two opposite forces are at work. One the one hand, the diminishing return of R&D to the number of product lines generates higher growth for small firms when financial constraints are slack. On the other, firms with low productivity, which also makes them small in size, find it more difficult to conduct R&D, for R&D expenditure is increasing at the aggregate level of productivity. Which force dominates depends on the tightness of financial constraints.

The model is also consistent with three observations from the data. Smaller firm R&D growth rates drop more precipitously during the Great Recession. Financially constrained firms switch to internal innovation, measured by a higher share of internal patents. In addition, the relation between size and growth rate for U.S. public innovative firms change from negative to statistically insignificant after 2007. The latter two are analyzed by panel regression, and are robust to various econometric models, variable definitions, identification strategies and sample selection.

Due to the complexity of the equilibrium, the model is solved using reinforced learning techniques. In particular, the state space of the value function is expanded to include firm-specific state variables, aggregate endogenous variables, to-be-estimated parameters and discretized aggregate productivity distribution. Then I approximate the value function by a neural network.

The financial constraint can tighten because of either reduction in profit pledgeability or drop in aggregate demand. Accordingly, I test three adverse scenarios: a 5% drop in pledgeability, a 5% drop in aggregate demand or both. The second case shows a larger decrease in growth rate and welfare, while the third one matches the empirical growth rate and size-growth relation, suggesting that the Great Recession is likely a combination of financial and demand shocks.

Lastly, I consider two R&D policies. In one every firm receives the same rate of R&D subsidy (uniform) while in the other small firms obtain higher subsidies, but both policies have the same average subsidy rate of 10%. The size-dependent policy yields a higher growth rate and higher welfare improvement.

There are several extensions possible for future research. First, I consider only permanent change in model parameters. Since the Great Recession left a long-lasting impact on the economy and firms were unclear of the persistence of shocks at the time, it is not inappropriate to simplify the analysis and consider a permanent shift, but it will be interesting to study the transition dynamics as well. Second, if firm-level data are available, one can also examine how financial development affects the growth of small firms in a cross-country setting. Lastly, as in [AAA17], a more comprehensive analysis should incorporate fixed operating costs, resources (mis)allocation and more complex R&D policies.

CHAPTER 3

Bank Profitability and Financial Stability

Jointly with with TengTeng Xu, and Udaibir S. Das[1]

3.1 Introduction

The Global Financial Crisis (GFC) of 2007-2009 and the ensuing period of low interest rates have renewed interest among policy makers on the importance of bank profitability for financial stability. Despite the subsequent recovery, the return on equity of many banks remains below the cost of equity. With valuations below the balance sheet value of banks[2], the market's assessment of banks' ability to overcome profitability challenges is not optimistic.

The existing literature on bank profitability and its impact on financial stability reports mixed evidence. First, on profitability and risks, some researchers found that higher profitability leads to higher charter value (i.e., long-term expected profitability) and therefore less risk-taking by banks ([Kee90, BKT09]). Others suggest that high profitability could loosen leverage constraints and lead to more risk-taking ([MRV16]). Furthermore, high profits in good times could be an indicator of systemic tail risk in bad times ([MNP18]). Second, there is mixed evidence on the impact of non-interest income (NII) on risks ([BDV07, EHH10]). More recently, some researchers found that the impact on financial stability depends on the type of non-interest income ([Koh14, DT13]).

---

[1]IMF Working Paper No. 19/5. See link `https://www.imf.org/en/Publications/WP/Issues/2019/01/11/Bank-Profitability-and-Financial-Stability-46470`. We are grateful to John Caparusso, Martin Cihak, Ehsan Ebrahimy, Javier Hamann, Fei Han, Mindaugas Leika, Graeme Littler, Hiroko Oura, Yizhi Xu, and the participants of the IMF's MCM Quantum Seminar for useful feedback. All remaining errors are our own.

[2]The low profitability of banks has been highlighted recently in the IMF's Global Financial Stability Reports (October 2016, 2017) and IMF led Financial Sector Assessment Programs (Euro Area 2018, Spain 2017, Japan 2017, Germany 2016, and Ireland 2016).

Motivated by such mixed evidence, this paper investigates the theoretical and empirical relationships between bank profitability and financial stability, taking into account bank business models (e.g., retail vs. wholesale orientation) and different types of NII activities. In this regard, we analyze not only the link between the *level* of bank profitability and financial stability, but also the deeper question of how the *source* of bank profitability affects financial stability. Several measures of bank business models and characteristics shed light on the *source* of bank profitability. For example: the NII share and the loan-to-asset (LTA) ratio provide insights on banks' reliance on NII and non-traditional business or activities; the deposit-to-liability ratio captures the extent to which banks rely on wholesale funding to cut costs on the liability side of the balance sheet[3]; the leverage ratio in part reflects banks' risk-taking behavior and the risks undertaken by banks to generate income; and competition measures such as the Lerner index of a firm's market power captures the extent to which banks rely on mark-up and market power to make profits.

This paper begins by setting out a stylized theoretical model that underpins the analytical relationship between bank profitability and financial stability by explicitly capturing the role of NII and retail-oriented business models. In this model, banks choose the amount of retail-oriented and market-oriented NII activities[4] to maximize expected equity values, given the risk profile of these activities. The key mechanism in our theoretical model is the complementarity between retail-oriented NII activities and bank lending. When the LTA ratio is high, banks are more inclined to engage in retail-oriented NII activities given the existing retail client base. On the other hand, when the LTA ratio is low, banks may choose to engage in market-oriented NII activities which tend to be riskier and providing limited diversification benefit from a financial stability prospective.

The theoretical model predicts that idiosyncratic risk, defined as the value-at-risk (VaR) of equity and the expected default frequency (EDF) proxy, decreases as both

[3]While wholesale funding provided cost savings prior to the crisis, there is some evidence that it became more expensive than retail funding after the crisis in some countries.

[4]Market-oriented business lines include underwriting, trade execution commissions, and investment-banking service. Retail-oriented business includes payment services fees, insurance commissions, and fiduciary income. The risk and return profile depend on the specific NII activity. For a summary of stylized facts, see [Sti02].

short-term book profitability (i.e., return on average assets, or ROAA) and long-term expected profitability (i.e., charter value) rise. Profits reduce risks by providing equity buffers and encouraging prudence, thereby reducing risk-taking. In addition, when the LTA ratio is below a certain threshold, idiosyncratic risk increases as the NII share[5] rises. We derive testable hypotheses on the relationship among bank profitability, business models, and financial stability.

In the empirical analysis, we apply dynamic panel regression approaches to examine the determinants of financial stability and profitability, and test the hypotheses derived from the theoretical model. Financial stability is captured by both idiosyncratic and systemic risk measures. Idiosyncratic risk is measured by market-based risk measures, including the historical VaR of equity prices and Moody's EDF, while the contribution to systemic risk is measured by the delta CoVaR ([AB16]). Bank profitability is measured by ROAA, ROAE (return on average equity), risk-adjusted returns, and the price-to-book ratio (a proxy for charter value). In our analysis, we not only control for business model measures, but also more generally, bank characteristics, structural and cyclical conditions, as well as monetary and fiscal policy variables. In our empirical analysis, we examine the "average" relationship between bank profitability, business models, and financial stability from 2004 to 2017, capturing both crisis and normal times. In this sense, our analysis is more general compared with papers that focus on crisis episodes alone. We focus our attention on 431 publicly traded banks as we capture market-based measures of bank profitability and financial stability.

Empirical results reveal several important interactions among bank profitability, business models, and financial stability, and confirm the hypotheses from the theoretical model. First, profitability (ROAA) and the price-to-book ratio are negatively associated with both the contribution to systemic risk ($\Delta CoVaR$) and idiosyncratic risks measured by the VaR (95 percent[6]) and the EDF of banks. Second, a high NII share tends to be associated with higher idiosyncratic and contribution to systemic risks when the LTA is low (i.e., when a bank's business model is less retail-oriented), as predicted by our theoretical model. Third, low competition is associated with lower idiosyncratic risk but

[5]The NII share is defined as the ratio of NII to operating income.

[6]We define VaR as the 95% quantile of "loss", which is the inverse of rate of return.

higher contribution to systemic risk. In addition, our results confirm that high leverage and an over-reliance on wholesale funding are associated with higher idiosyncratic and contribution to systemic risks.

This paper contributes to the existing literature on bank profitability and financial stability from both theoretical and empirical perspectives[7]. Theoretically, this paper provides one of the first models to pin down the analytical relationship between risks and bank profitability, accounting for the interaction between NII and retail-oriented business models. Most papers focus on a narrow set of NII activities, especially market-oriented ones such as securitization and trading (e.g., [SV10, BR16]), yet fee-based traditional retail-oriented business is another crucial component of NII. In our model, we consider both retail-based and market-based NII and the distinction is general in nature: only the former is complementary with respect to bank lending. We then derive explicitly the impact of bank profitability, the NII share, and the LTA ratio on idiosyncratic risks measured by VaR and EDF.

The empirical contribution of our paper is three-fold. First, the paper is one of the first comprehensive empirical analysis on the determinants of bank idiosyncratic risks and their contribution to systemic risks, accounting for bank profitability, business models, structural and cyclical conditions, and policy responses during recent crises. Second, we contribute to the empirical literature on NII and their financial stability implications by explicitly controlling for retail vs. wholesale business models in a cross-country setting without reliance on confidential supervisory data. Earlier papers either used detailed supervisory data or categorize NII according to local accounting standards in country-specific studies (e.g., [Koh14, DT13]). In this paper, using an interaction term between share of NII and the LTA ratio, we are able to control for the type of NII activities and bank business models in a cross-country setting, even when accounting standards differ across countries[8]. Third, we examine the relationship between the forward-looking measure of risks (EDF) and bank profitability empirically. Earlier literature has largely

[7] A literature survey on the determinants of risks and bank profitability can be found in Appendix 5.1.

[8] Furthermore, publicly available data sources such as the S&P Global Market Intelligence's SNL database and Fitch Connect do not provide further breakdown on the type of NII and loans in a consistent manner across countries, in part, due to different accounting standards and reporting requirements among countries.

focused on backward-looking measures ([DT13]).

Finally, our paper contributes to policy discussions on the role of bank profitability for financial stability. The results demonstrate that the source and the sustainability of bank profitability has important financial stability implications, as an over-reliance on market-based NII activities, leverage, and wholesale funding is associated with higher idiosyncratic risk and contribution to systemic risk. Furthermore, the impact of bank consolidation on competition should be addressed in policy discussions, as low competition is associated with a high contribution to systemic risk.

The rest of the paper is structured as follows. We first present a stylized theoretical model on the relationship between profitability and financial stability in Section 3.2. Section 3.3 presents the data, stylized facts, and the empirical methodology. We then discuss the empirical findings on the determinants of risk and bank profitability in Section 3.4. Finally, we offer some concluding remarks and discuss policy implications in Section 3.5.

3.2 Bank Profitability and Risks: A Stylized Theoretical Model

To anchor analytical relationships between bank profitability and financial stability, we outline below a stylized model accounting for bank business models. The focus of the model is to capture both retailed-based and market-based NII activities, and the non-linear impact of NII on banking risks. To keep it tractable and focused, we abstract from modelling a dynamic programming problem, as it is not essential for capturing the stylized relationships among bank profitability, business models, and financial stability.

3.2.1 Model Setup

Bank Balance Sheet

In the stylized theoretical model, we consider a static setting of a representative risk-neutral bank with the following balance sheet structure:

The balance sheet constraint is given by

$$L + N_r + N_m = D + E \equiv A \tag{3.1}$$

Assets	Liabilities
L	D
N_r	E
N_m	

where N_m stands for the assets related to market-based NII activities, such as underwriting, trade commissions, and investment-banking services, and N_r captures retail-based NII activities, such as payment services fees, insurance commissions, lending service fees, and fiduciary income. N_r and N_m are the assets devoted to NII activities at the beginning of the period[9]. L represents loans, D deposits, E equity, and A bank assets. For simplicity and tractability, we assume D, L and E to be exogenous, and that the capital constraint is binding, $E = eA$, where e is the reciprocal of the leverage ratio. The assumption that L is exogenous is not unreasonable in our stylized model, as L can be regarded as a proxy for the *retail customer base of a bank* and typically cannot change quickly. Similar reasoning is applied to equity E by assuming some equity issuance costs or frictions.

Bank Profit Function and Shocks

The bank's profit function is given as

$$\tilde{\Pi} = (1-x)\tilde{r}_L L + \tilde{r}_m N_m + \tilde{r}_r N_r^{\alpha} L^{1-\alpha} - c_m N_m - c_r N_r - c_f A - r \qquad (3.2)$$

where tilde notation denotes random variables, and returns are normally distributed as $\tilde{r}_i \sim N(r_i, \sigma_i^2)$ with $i = L, m, r$. For simplicity, we assume $\tilde{r}_i$ are mutually independent[10]. c_m, c_r, and c_f are cost parameters where $c_m < r_m$. r_D is the deposit rate and x denotes the problem loan ratio. The deposit rate r_D can be viewed as funding cost in our stylized model.

A key structure in the model is the Cobb-Douglas production function $N_r^{\alpha} L^{1-\alpha}$ of

[9]For example, if retail-based NII includes payment service fees, then N_r represents the payment network or system's assets (e.g., ATMs, software, machinery).

[10]This assumption is not important and will not alter the main results. For details, please refer to the discussions after proposition 2.

retail-based NII activities[11]. It ensures homogeneity of degree one with respect to inputs L and N_r, as well as the complementarity between retail-based lending business L, and retail-based NII activities N_r. The complementarity is motivated by the fact that most retail-based NII activities share the same customer base (and some employee skills) as the lending business.

Bank Objective Function

The bank's objective function is given by

$$\max_{N_m, N_r} \mathbf{E}(\tilde{\Pi}) + E \tag{3.3}$$

subject to the balance sheet constraint $L + N_r + N_m = D + E$. Note that the bank's survival probability is given by $q = Prob(\tilde{\Pi} + E \geq 0)$. Following [MV96], we abstract from the assumption of limited liability.

We normalize the bank's objective function by bank asset A and take expectations. The normalized objective function is then given by

$$\max_{n_m, n_r} \mu_\pi + e \tag{3.4}$$

subject to $l + n_r + n_m = 1$, where $\mu_\pi \equiv \frac{\mathbf{E}(\tilde{\Pi})}{E} = (1-x)r_L l + r_m n_m + r_r n_r^\alpha l^{1-\alpha} - c_m n_m - c_r n_r - c_f - r_D(1-e)$, $e = E/A$, $l = L/A$, $n_M = N_M/A$ and $n_r = N_r/A$.

Note that n_m and n_r capture market-based and retail-based NII intensity (share of NII activities in bank asset, different from income), l is the LTA ratio, and μ_π captures the *expected return on asset* (ROA).

[11] The specific functional forms for the returns on retail-oriented NII (N_r) and market-oriented NII (N_m) are not critical for the theoretical results. Instead of the linear specification of the return on N_m, one can also assume a Cobb-Douglas production function, $N_m^\beta L^{1-\beta}$, where $\beta > \alpha$ for the market-based NII activities. The underlying reason for $\beta > \alpha$ is that retail-oriented NII is expected to have more complementarity with bank lending than with market-oriented NII activities. In other words, there could be complementarities between market-based NII activities and bank lending, but the degree of complementarity between retail-based NII activities and bank lending is expected to be higher than that of market-based ones.

Definition of Risks

As we are interested in the relationships between bank profitability and financial stability, we focus on two types of risks that are particularly relevant for financial stability considerations. First, we consider the default probability of a bank, measured by its overall credit risk or solvency. Second, we are interested in the tail risks faced by a bank. Based on our stylized theoretical model, we define the EDF proxy (default probability) and the VaR of individual banks as follows:

Expected Default Frequency Proxy (EDF)

$$EDF \equiv 1 - q = Prob(\tilde{\pi} + e < 0). \tag{3.5}$$

The EDF proxy is defined as one minus the survival probability of the bank[12]. A bank defaults in our model when equity is below zero.

Value-at-Risk (VaR)

$$Prob(|loss| \geq VaR) = Prob(-\tilde{\pi} - e \geq VaR) = 0.05. \tag{3.6}$$

The VaR is defined as the 95 percentile of equity loss in this model, where higher VaR signifies higher tail risks.

3.2.2 Solutions and Propositions

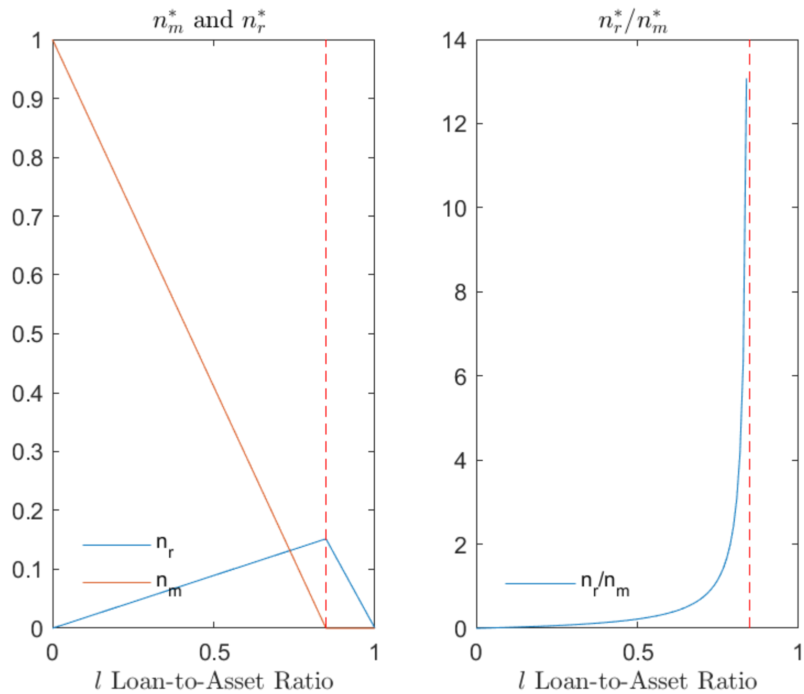
We solve the bank's optimization problem by taking first order conditions with respect to n_m and n_r, subject to its budget constraint. The resulting first order conditions are given as follows:

$$\begin{aligned} &[n_m] : r_m = c_m + \varphi \text{ if } n_m > 0 \\ &[n_r] : \alpha r_r n_r^{\alpha-1} l^{1-\alpha} = c_r + \varphi \text{ if } n_m > 0 \end{aligned} \tag{3.7}$$

where φ is the Lagrange multiplier of $l + n_r + n_m = 1$.

[12]This definition of the EDF proxy is applicable to a more general concept of default probability.

Figure 3.1: Optimal Market-based and Retail-based NII Intensity



We then rewrite the first order conditions with superscript $*$ denoting the optimal value of choice variables:

$$\begin{aligned} &\text{if } l \geq \frac{1}{1+k} : n_r^* = 1 - l, n_m^* = 0 \text{ and } n_r^*/n_m^* = \infty \\ &\text{if } l < \frac{1}{1+k} : n_r^* = kl, n_m^* = 1 - l - kl \text{ and } n_r^*/n_m^* = \frac{kl}{1 - l - kl} \end{aligned} \tag{3.8}$$

where $k = \left(\frac{\alpha r_r}{c_r + r_m - c_m}\right)^{\frac{1}{1-\alpha}} > 0$.

If $l < 1/(1+k)$, the first order conditions imply an interior solution where the optimal retail-based NII intensity n_r^* is a positive function of the LTA ratio l, reflecting the complementarity between n_r and l. If $l1/(1+k)$, the first order conditions imply a corner solution where the optimal market-based NII intensity n_m^* is equal to zero (Figure 3.1, left panel).

Risks and Profitability

Based on the model solutions, we can derive Proposition 1 on the relationship between bank risks and profitability.

Proposition 3. *Bank idiosyncratic risks measured by EDF and VaR are decreasing in the (expected) ROA* μ_π^*.

$$\frac{\partial EDF}{\partial \mu_\pi^*} < 0, \frac{\partial VaR}{\partial \mu_\pi^*} < 0$$

Proof. See Appendix 5.2. ■

The intuition for the negative relationship between idiosyncratic risks and bank profitability is that per-period profit μ_π^* (or the book value of profit) provides a buffer against negative shocks to bank capital. Higher μ_π^* means larger buffers and reduced default risk, which lowers idiosyncratic risks[13].

Risks, NII, and the LTA Ratio

Having established the analytical relationship between bank profitability and risks, we next examine the source of bank profitability and the relationship to bank risks. We are particularly interested in the role of NII activities for risks, accounting for bank business models. From the first order conditions of the model, we derive the Lemma 1 below.

Lemma 4. *The ratio of retail-based NII intensity to market-based NII intensity is increasing in the LTA ratio:*

$$\frac{\partial (n_r^*/n_m^*)}{\partial l} \geq 0.$$

Proof. From the solution to first order conditions, we have

$$\begin{aligned} &\text{if } l \geq \frac{1}{1+k} : \frac{\partial (n_r^*/n_m^*)}{\partial l} = 0 \\ &\text{if } l < \frac{1}{1+k} : \frac{\partial (n_r^*/n_m^*)}{\partial l} = \frac{k}{(1-l-kl)^2} > 0. \end{aligned}$$

■

[13]Some papers that internalize borrowers' decisions argue that if lower profitability is a result of lower interest rate margins, then reduced credit rationing in the loan market will improve the average quality of loan applicants, which ultimately translates to lower bank risks ([BD05, SW81]). Given that interest rates were very low in our sample period of 2004 to 2017, it is reasonable to abstract from the credit rationing channel. Instead, we focus on the equity buffer channel and the charter value channel of bank profitability in our stylized theoretical model.

Lemma 4 states that the composition of NII will change with respect to the LTA ratio. The result is also shown in the right panel in the Figure 3.1[14]. Intuitively, this result follows from the complementarity between n_r and l, because of the term $n_r^\alpha l^{1-\alpha}$ in expected profitability μ_π. The LTA ratio l is a proxy of the retail business for a bank. Higher l is associated with more retail clients, which makes developing retail-based NII activities easier (the marginal benefit of n_r depicted by $\alpha r_r n_r^{\alpha-1} l^{1-\alpha}$ increases in l). So a high-l bank willingly leans toward more retail-based NII, and n_r^*/n_m^* increases.

Denote the (expected) share of the (overall) NII as

$$s = \frac{NII}{NII + II} = \frac{r_m n_m^* + r_r n_r^{*\alpha} l^{1-\alpha}}{r_m n_m^* + r_r n_r^{*\alpha} l^{1-\alpha} + (1-x) r_L l}. \tag{3.9}$$

Under the interior solution, when $l < 1/(1+k)$, the share of NII s can be rewritten as

$$s = \frac{r_m - c_m - (1+k)l(r_m - c_m) + (rk^\alpha - c_r k)l}{r_m - c_m - (1+k)l(r_m - c_m) + (rk^\alpha - c_r k)l + (1-x)r_L l}$$

and it is straight forward to show that

$$\frac{\partial s}{\partial n} = \frac{\partial s}{\partial (n_r + n_m)} = \frac{\partial s}{\partial (1-l)} = -\frac{\partial s}{\partial l} > 0.$$

In other words, the expected NII share is increasing in total NII intensity $n_r + n_m$. This leads to the following proposition regarding NII activities and bank risks.

Proposition 4. *When LTA ratio (l) is below a certain threshold ($\underline{l}$), higher NII share (s) will lead to higher VaR and EDF:*

$$\frac{\partial EDF}{\partial s} > 0 \textit{ and } \frac{\partial VaR}{\partial s} > 0 \textit{ if } l < \underline{l} = \frac{(1+k)\sigma_m^2}{(1-x)^2\sigma_L^2 + (1+k)^2\sigma_m^2 + k^{2\alpha}\sigma_r^2} < \frac{1}{1+k}$$

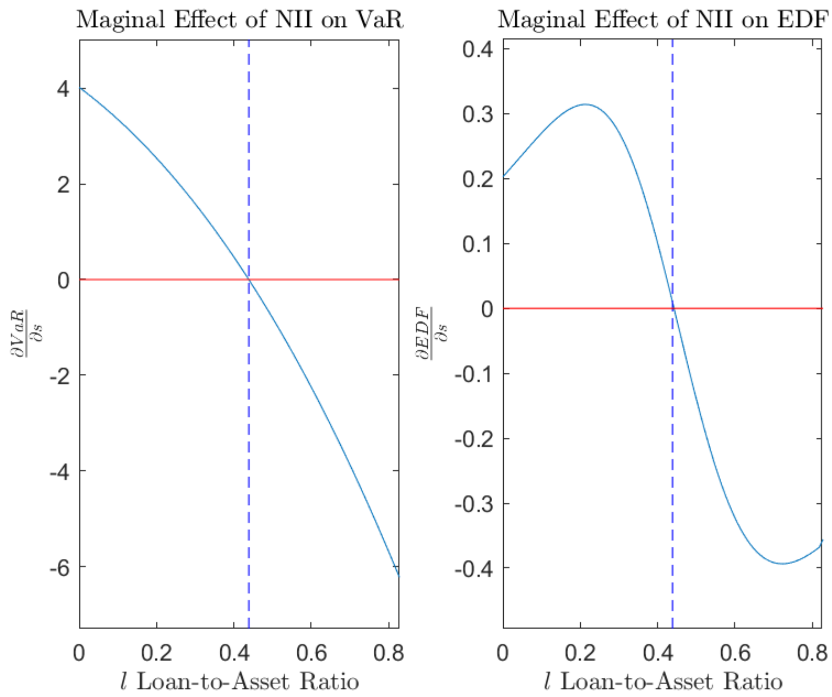
under a regularity condition[15].

Proof. See Appendix 5.2. ■

[14]In Figure 3.1, at the red dotted line $l = 1/(1+k)$, the bank's optimization problem yields a corner solution as $n_m^* = 0$.

[15]*The regularity condition is that the problem loan ratio* $x < 1 + \frac{(r_m - c_m)(\frac{1-\alpha}{\alpha}k - 1) + \frac{1-\alpha}{\alpha}kc_r}{r_L}$. *This parameter assumption is reasonable, as the average value of x observed empirically in our sample is less than 5%.*

Figure 3.2: Non-interest Income Share and Risks



The effect of the NII share on idiosyncratic risks (VaR and EDF) are illustrated in Figure 3.2. The dotted blue line denotes $\underline{l}$. When $l \leq \underline{l}$, the partial derivatives are positive, meaning that increasing NII share will result in higher idiosyncratic risks. This is because, as noted in Lemma 4, n_r^*/n_m^* decreases as l declines.

To understand the underlying mechanism, note that bank assets are a portfolio with three sources of return: loans (r_L), market-based NII activities (r_m), and retail-based NII activities (r_r). The overall bank idiosyncratic risk is a function of the portfolio weights l, n_r, and n_m. As long as not all "assets" are perfectly correlated, there is benefit from diversification (i.e., it is risk-reducing to participate in NII activities $n_m, n_r > 0$). An over-reliance on any particular type of assets is sub-optimal.

The interesting result is that *the profit-maximizing weights* (n_m^*, n_r^*) *chosen by banks are not necessarily risk-minimizing*, even if in this model, higher expected profits *do* help reduce risks (i.e., Proposition 3). Crucially dependent on a bank's *business model* (differentiated via loan-to-asset ratio l), it is preferable to increase or decrease the share of NII activities $n_m^* + n_r^*$ (or equivalently change l) from a financial stability/risk-control perspective. When l is too low ($l \leq \underline{l}$), further action to weigh in NII activities (reduce l) will increase risks ($\frac{\partial EDF}{\partial s} > 0, \frac{\partial VaR}{\partial s} > 0$). A bank too reliant on loan income ($l > \underline{l}$)

should do the reverse and increase NII (i.e., to the right of the dotted line in Figure 3.2) to capture their diversification benefits.

Riskiness of Market-Based NII Activities n_m

So far I have abstracted from the discussion of σ_m^2 and σ_r^2 (variance of the returns from market-based and retailed-based NII activities). What if $\sigma_m^2 > \sigma_r^2, \sigma_L^2$, as is empirically true ([Sti02])?

From $\underline{l} = \frac{(1+k)\sigma_m^2}{(1-x)^2\sigma_L^2+(1+k)^2\sigma_m^2+k^{2\alpha}\sigma_r^2}$, we can see $\underline{l}$ is increasing in σ_m^2. As market-based NII become more volatile, *more* banks will fall into the range where $l < \underline{l}$ and they become over-reliant on riskier market-based NII activities (n_m is too high, or equivalently n_r/n_m too low). For such banks, it would be advisable to reduce the exposure to NII activities and increase interest income share since $\frac{\partial EDF}{\partial s} > 0$, $\frac{\partial VaR}{\partial s} > 0$. Only for banks with high loan-to-asset ratio ($l > \underline{l}$) is there risk-reducing benefit from increasing NII activities, because they will incline toward safer retail-based NII activities n_r (endogenously determined by the complementarity between n_r and l).

In the most extreme case when $\sigma_m^2 \to \infty$, we have $\underline{l} \to \frac{1}{1+k}$. So only banks with loan-to-asset ratio higher than $\frac{1}{1+k}$ should venture into NII activities from the financial stability standpoint, because only those banks have $n_m^* = 0$ (see Equation 3.8).

Retail-Based NII Activities n_r and Loan Income

What is the relationship between retail-based NII activities n_r and loan l? As a result of the common consumer based (which also gives rise to the complementarity) between n_r and l, it is reasonable to allow for some positive correlation between $\tilde{r}_r$ and $\tilde{r}_L$. For example, if $\tilde{r}_r$ and $\tilde{r}_L$ follow a bivariate normal distribution with correlation coefficient ρ, $\underline{l}$ becomes

$$\underline{l} = \frac{(1+k)\sigma_m^2}{(1-x)^2\sigma_L^2 + (1+k)^2\sigma_m^2 + k^{2\alpha}\sigma_r^2 + \alpha(1-x)k^{\alpha}\rho\sigma_L\sigma_r}$$

All the aforementioned results will still hold in this more general case.

Our theoretical finding is consistent with that of [MRV16] and [MNP18], which suggest that NII activities could lead to higher risks. The former propose that higher profitability in core business relaxes a bank's leverage constraint, enabling more risk-taking in non-core (i.e., NII) business. The latter also argue that NII is the main culprit of highly volatile returns (high profit in good times and high loss in bad times). Given that NII is a broad catogory, with distint risks for different NII components ([Sti02]), our model endogenizes a bank's decision on which NII activities to focus on, which in turn draws different predictions on the effect of NII activities on bank risks.

Discussion of Systemic Risks

Besides idiosyncratic risks, the activities of a financial institution can contribute to systemic risk (i.e., to the overall financial system). In this model, market-based NII activities n_m can have implications on a bank's systemic risks, even if we do not model the financial system (and its risk) explicitly. First, participation in market-making activities and securitization (among other examples of n_m) will increase the risks a bank imposes on the financial system (e.g., Lehman Brothers). Second, as returns of n_m, $\tilde{r}_m$, can be highly correlated to the market and across banks, n_m can increase the correlation and contagion risks, contributing to higher systemic risks [BCH17]. In other words, a higher value of n_m is then expected to contribute to higher systemic risks.

Because this is a model focused on bank idiosyncratic risks, we cannot directly map our theoretical results to systemic risk analysis. Still, it is interesting to see empirically whether the level and source of profits will have similar effects on bank systemic risks, which we will come back to in Section 3.4.

3.2.3 Extensions with Bank Charter Value

In addition to the book value of profitability (or per period profit), another common measure of profitability is the price-to-book ratio, which can be interpreted as the charter value of a bank, or expected discounted future profits. A high charter value can have a disciplinary effect on bank risk-taking behavior. Motivated by this consideration, we extend our baseline model to include the interaction of bank charter value and idiosyncratic risks. For analytical simplicity, we consider a case where a bank has already made the optimal choice on NII activities (i.e., $n_m = n_m^*$ and $n_r = n_r^*$) and isolate the implication of charter value on banking risks alone.

In the extended model, the bank is also subject to a random shock $-zA$ to equity[16], where z follows a Bernoulli Distribution:

$$z = \begin{cases} \epsilon & \text{with probability } 1-p \\ 0 & \text{with probability } p. \end{cases}$$

One interpretation of shock z is an operational risk shock. The likelihood that the bank is affected by the random shock z depends on the intensity of its monitoring. The more intense the monitoring activity (high monitoring cost), the lower the likelihood that it will be affected by shocks to equity. The monitoring or risk management cost is given by:

$$C(p) = -\frac{1}{2}bp^2A$$

where $C(p)$ is a function of asset size A, the probability of the shock p, and a constant b. A banks is therefore incentivized to monitor—in order to reduce the expected equity impact from random shock—as long as the marginal monitoring cost does not exceed the marginal impact on bank equity from the random shock.

Also, let V denote the continuation value of bank equity. Other interpretations of V can be the charter value, discounted future profits, or the market value of equity (see [FR08], Chapter 3.5). For tractability, we assume that V is exogenously given[17].

[16] Therefore, there are two sources of randomness in the extended model. The first one is the randomness to asset returns $\tilde{r}_L, \tilde{r}_m$ and $\tilde{r}_r$, which follow a normal distribution as explained earlier. At optimal levels of n_m and n_r, the expected returns are fixed while the actual returns remain random. The second source of randomness is a shock z to bank equity. In the extension, a bank's only choice variable is the probability of the equity shock z.

[17] In this stylized model, we do not endogenize the continuation value of equity V, as it is not crucial for the derivation of the analytical relationships between bank profitability and financial stability. One could potentially extend the model to a dynamic setting where V will depend on the entry cost of banks.

New Objective Function

The bank's new objective function[18] is then given by:

$$\max_p \mathbf{E}(\tilde{\Pi} + E - zA) - \frac{1}{2}bp^2 A + qV \tag{3.10}$$

subject to the balance sheet constraint $L + N_r + N_m = D + E$. We then normalize the bank's objective function by bank asset A and take expectations. The normalized objective function is then given by

$$\max_p \mu_\pi^* + e - \epsilon(1-p) - \frac{1}{2}bp^2 + qev \tag{3.11}$$

where $v = \frac{V}{E}$ = Price-to-Book Ratio and $e = E/A$ captures the inverse of leverage.

Since the bank is subject to a new equity shock, the bank's survival profitability is modified to q', reflecting the Bernoulli Distribution of shock z:

$$\begin{aligned} q' =& Prob(\tilde{\pi} + e - z \geq 0) \\ =& p \cdot Prob(\tilde{\pi} + e - z \geq 0) + (1-p) \cdot \tilde{\pi} + e - z \geq \epsilon) \\ =& p\left[\Phi(\frac{\mu_\pi + e}{\sigma_\pi}) - \Phi(\frac{\mu_\pi + e - \epsilon}{\sigma_\pi})\right] + \Phi(\frac{\mu_\pi + e - \epsilon}{\sigma_\pi}). \end{aligned}$$

Bank idiosyncratic risk measures can also be modified to account for the new equity shock:

$$\begin{aligned} &EDF' = \equiv 1 - q' = Prob(\tilde{\pi} + e - z \geq 0) \\ &Prob(|Loss| \geq VaR) = Prob(-\tilde{\pi} - e + z \geq VaR') = 0.05. \end{aligned}$$

The optimal shock probability p^* is then given by the first order condition with respect to p:

$$p^* = \frac{\epsilon + \left[\Phi(\frac{\mu_\pi^* + e}{\sigma_\pi^*}) - \Phi(\frac{\mu_\pi^* + e - \epsilon}{\sigma_\pi^*})\right] ev}{b}. \tag{3.12}$$

[18] With limited liability, there is a threshold $\hat{V}$ (charter value) below which banks engages in risk-taking behavior, as the low charter value is not sufficient to discipline them ([FR08], Chapter 3.5). This mechanism is not our focus, as the empirical evidence is largely in favor of the mechanism that charter value defers risk taking (see, for example, [Kee90] [BKT09]), which is captured in our modeling framework.

It is interesting to note that p^* is positively related to the price-to-book ratio v, and the inverse of leverage e. Recall that p^* is the probability that the equity impact of shock z is zero (or minimum), and a higher p^* is associated with more intense monitoring or higher monitoring costs. One interpretation is that rising price-to-book values or falling leverage (higher equity) incentivizes banks to monitor and to reduce the equity impact of shocks. Based on these consideration, we derive two propositions that underpin the analytical relationships between idiosyncratic risks and the price-to-book ratio, and between idiosyncratic risks and bank leverage.

Risks and the Price-to-Book Ratio

Proposition 5. *Bank idiosyncratic risks measured by EDF' and VaR' are decreasing in the price-to-book ratio v.*

$$\frac{\partial EDF'}{\partial v} < 0 \ , \ \frac{\partial VaR'}{\partial v} < 0$$

Proof. See Appendix 5.2 ■

The intuition for the negative relationship between bank idiosyncratic risks and the price-to-book value, v, is that higher charter value or long-term profits (captured by v) deters risk-taking behavior of banks. A bank is only able to retain its charter value if it survives at the end of the period. Therefore, the higher the v, the higher the incentive for banks to reduce risk-taking and avoid potential bankruptcy. This finding is consistent with that of [Kee90] and subsequent papers (e.g., [BT10, MV00, Rep04]) that charter value provides incentive for prudence[19].

[19]Some papers that internalize borrowers' decisions argue that if lower profitability results from lower interest rate margins, reducing credit rationing in the loan market will improve the average quality of loan applicants, which ultimately translates to lower bank risks ([BN05, SW81]). Given that interest rates were very low in our 2004-2017 sample period, it is reasonable to abstract from the credit rationing channel. Instead, we focus on the equity buffer channel and the charter value channel of bank profitability in our stylized theoretical model.

Risks and Leverage

We can also derive that a higher equity to asset ratio, or lower leverage, will reduce bank idiosyncratic risks.

Proposition 6. *Bank idiosyncratic risks measured by EDF' and VaR' are decreasing in e (increasing in leverage* $1/e$*).*

$$\frac{\partial EDF'}{\partial e} < 0 \ , \ \frac{\partial VaR'}{\partial e} < 0$$

Proof. See Appendix 5.2 ■

Higher equity-to-asset ratio e implies more "skin in the game for banks, and thus they will have higher incentives to monitor and reduce risk-taking behaviors to avoid defaults. As discussed earlier, this is reflected by the fact that p, the probability of no equity shock (a choice variable to the bank), is negatively related to e, as in equation 3.12. Additionally, higher equity increases the buffer against negative shocks for banks, which reduce bank risks mechanically through accounting relationships in bank balance sheets[20].

3.3 Stylized Facts and Empirical Methodology

3.3.1 Key Variables

We consider a sample of 431 publicly-traded banks in our empirical analysis. The sample includes all public banks in the U.S. and developed Europe, and all other Global Systemically Important Banks (GSIBs)[21]. The sample period spans from 2004 to 2017 and the

[20] It should be noted that the relationship between idiosyncratic risk and leverage also holds in the simple baseline model. The intuition is that higher equity leads to higher buffers against negative shocks. In the extended model, the channels through which leverage affects idiosyncratic risks are richer, as they not only pertain to equity buffers, but also to the charter value of banks and bank incentives to monitor. For completion, we also provide a proof on the negative relationship between idiosyncratic risk and leverage in the simple baseline model (Corollary 6.1) in the Appendix 5.2.

[21] Among the publicly traded banks, 308 are from the U.S. and 115 are from developed Europe. In addition, the sample includes the eight GSIBs from outside the U.S. and the Europe. The average asset size for U.S. banks, European banks, and GSIBs are \$53 billion, \$274 billion, and \$1710 billion, respectively. The GSIB list follows the classification by the Financial Stability Board (FSB) in 2017. `http://www.fsb.org/2017/11/2017-list-of-global-systemically-important-banks-g-sibs/`

data source is S&P Global Market Intelligence's SNL database[22].

Profitability Measures

Five profitability measures are considered in the empirical analysis: ROAA, ROAE, risk-adjusted ROAA, risk-adjusted ROAE, and the price-to-book ratio[23]. The risk-adjusted profitability measures are computed as the ratio of headline profitability measures (ROAA or ROAE) and their standard deviation for the sample period (2004 to 2017) for each bank. The price-to-book ratio is the ratio of the market value of equity (share price) and the book value of equity, which is often used as a proxy for expected profitability (or charter value).

Financial Stability Measures

Financial stability is captured by three systemic and idiosyncratic risk measures. Systemic risk is measured by the delta $\Delta CoVaR$ ([AB16]), while idiosyncratic risk is measured by the five percent VaR and Moody's EDF[24].

Idiosyncratic Risk Measures

Idiosyncratic risk is measured by two market-based risk measures: historical VaR based on annualized daily equity return at 5 percent and Moody's EDF.

The five percent VaR is computed as the lowest five percent quantile of daily equity returns in a particular year. Moody's EDF is a forward-looking measure of actual probability of default of a bank over a specified period of time (one year in this application). According to the Moody's EDF model, a bank defaults when the market value of its assets falls below its liabilities[25].

---

[22]For sources and definitions of the variables, see Appendix 5.3.

[23]The choice of the profitability measures is motived by our theoretical model. ROAA and ROAE are empirical proxies for per-period profit (μ_π), and the price to book ratio is an empirical proxy for chart value or discounted future profits (V). In addition, we are interested in analyzing risk-adjusted ROAA and ROAE measures.

[24]Similarly, the choice of idiosyncratic risk measures is motivated by the theoretical model. The empirical VaR and EDF map directly to the risk measures (VaR and EDF) in the theoretical framework. We also consider an established systemic risk measure $\Delta CoVaR$ ([AB16]).

[25]See Moody's Analytics at `https://www.moodysanalytics.com` for details on constructing EDFs.

Systemic Risk Measure

Following [AB16], the $\Delta CoVaR$ is estimated using quantile regression on weekly data[26]:

$$X_t^i = \alpha_q^i + \gamma_q^i M_{t-1} + \epsilon_{q,t}^i$$

$$X_t^{System|i} = \alpha_q^{System|i} + \gamma_q^{System|i} M_{t-1} + \beta_q^{System|i} X_t^i + \epsilon_{q,t}^{System|i}$$

where X_t^i denotes the weekly equity return of bank i, $X_t^{System|i}$ the weekly system equity return conditional on bank i, and M_t the list of state variables. We use q to denote the qth quantile. We then use the predicted values from these regressions to obtain VaR and CoVaR conditional on state variables:

$$VaR_{q,t}^i = \alpha_q^i + \gamma_q^i M_{t-1}$$

$$CoVaR_{q,t}^i = \alpha_q^{System|i} + \gamma_q^{System|i} M_{t-1} + \beta_q^{System|i} VaR_{q,t}^i.$$

Finally, we compute $\Delta CoVaR_{q,t}^i$ for each bank as the difference between the qth percentile CoVaR and the median CoVaR:

$$\Delta CoVaR_{q,t}^i = CoVaR_{q,t}^i - CoVaR_{50,t}^i = \beta_q^{System|i}(VaR_{q,t}^i - VaR_{50,t}^i)$$

In this paper, we consider the 5th percentile $\Delta CoVaR$ in the empirical analysis. For European banks, Euro Stoxx bank returns were used to capture financial sector returns, while S&P 500 financial index returns were used for U.S. banks and other GSIBs.

We consider a set of state variables in the quantile regression estimations, including interest rates, term structure of interest rates, liquidity risk, credit risk, market returns, market volatility, and excess return of the financial sector over the real estate sector. For the most part, U.S.-specific state variables were used to construct the $\Delta CoVaR$ for U.S.

[26] For ease of illustration, we follow the same notation as in [AB16].

banks and other GSIBs[27], while Europe-specific state variables were applied to European banks. The exception was credit risk and VIX measures, which were used in both cases.

The following set of state variables were used in the estimation: (i) interest rates, measured by the change in 3-month German bond yields, and the change in 3-month T-bill rates; (ii) term structure of interest rates, measured by the change in the spread between 10-year and 3-month German government bond yields, and the change in the spread between 10-year and 3-month T-bill rates; (iii) liquidity risk, measured by the change in the difference between 3-month Euribor (Euro Interbank Offered Rate) and 3-month Germany bond yields, and the change in the difference between 3-month LIBOR and 3-month secondary market T-bill rates; (iv) credit risk, measured by the change in credit spreads between Moody's Baa-rated bonds and the 10-year Treasury rates; (v) market returns from the Euro Stoxx 50 and S&P 500 indices; (vi) market volatility, measured by the change in the VIX index; and (viii) excess return of the financial sector over the real estate sector, measured by the difference between the Euro Stoxx banks index returns and the MSCI Europe real estate index returns, and the difference between the S&P 500 financials index returns and the Dow Jones U.S. real estate index[28].

Bank Business Models and Characteristics

Bank business models are captured by four variables in the analysis. First, to measure the reliance of banks on NII we consider the share of NII to revenue. Second, the LTA ratio is used as a proxy for retail vs. wholesale business models for banks. Third, the deposit-to-liability ratio is used to capture banks' reliance on wholesale funding. Fourth, the asset-to-equity ratio is used to measure the extent of leverage. These four business model variables capture both returns and asset allocation of banks, and both the asset and the liabilities sides of bank balance sheets. The four variables are examined in detail

[27] Similar to [LMR12], we use the set of state variables sampled from the U.S. market as common conditional variables for other GSIBs.

[28] As a robustness check, we also consider a version of the $\Delta CoVaR$ estimation that controls for state variables based on world variables, in addition to regional ones. In this case, the financial sector return was measured by the return of the MSCI world financial index, the market return was captured by the MSCI world index return, and the excess return was measured by the difference between the MSCI world financial index returns and the MSCI world real estate index returns. The results are found to be very similar. For the rest of the paper, we focus on the $\Delta CoVaR$ analysis based on region-specific state variables.

in the analysis on the determinants of risks.

Several variables on bank characteristics are controlled for in the empirical analysis, including solvency, measured by the Tier 1 capital ratio; asset quality, captured by the problem loans ratio; and efficiency, captured by cost-to-income and cost of funds ratios. The cost-to-income ratio is measured by the ratio of operating expense to operating income and is a standard measure of operating efficiency ([BGH17, Int17]). The cost of funds ratio is captured by the interest incurred on liabilities as a percent of average noninterest-bearing deposits and interest-bearing liabilities.

In addition, the Lerner index was constructed to capture the market power or mark-up of banks. Following the specification in [BKT09], the Lerner Index for bank i was constructed as $L_{it} = (P_{it} - MC_{it})/P_{it}$, where P_{it} is the price of assets, measured by the ratio of total revenue to total assets, and MC_{it} is the marginal cost of total assets. The higher the value of the Lerner index, the easier it is for a bank to charge over its marginal costs, and therefore the greater its mark-up or market power.

The marginal cost of total assets for bank i at time t, MC_{it}, is computed as $MC_{it} = \frac{cost_{it}}{Q_{it}}[\beta_1 + \beta_2 \ln Q_{it} + \sum_{k=1}^{3} \phi_k \ln W_{k,it}]$, where $W_{1,it}$ is the ratio of personnel expense to total assets and a proxy for the input price of labor, $W_{2,it}$ is the ratio of interest expense to total deposits and a proxy for the input price of funds, and $W_{3,it}$ is the ratio of other operating and administrative expenses to total assets and captures the input price of fixed capital[29]. Q_{it} and $cost_{it}$ capture total assets and total costs, respectively. Furthermore, the coefficients β_1, β_2, and ϕ_k are estimated from the following cost equation:

$$\begin{aligned}\ln cost_{it} =& \beta_0 + \beta_1 \ln Q_{it} + \frac{\beta_2}{2} \ln Q_{it}^2 \\ &+ \sum_{k=1}^{3} \gamma_{kt} \ln W_{k,it} + \sum_{k=1}^{3} \phi_k \ln Q_{it} \ln W_{k,it} \\ &+ \sum_{k=1}^{3} \sum_{j=1}^{3} \ln W_{k,it} \ln W_{j,it} + \epsilon_{it}.\end{aligned}$$

As noted in the literature, the Lerner index has several advantages over alternative measures of market competition and concentration. First, the Lerner index can be com-

[29] Note that $W_{2,it}$ reflects market power in the deposit market. Alternatively, the Lerner index could be computed using the marginal cost estimation for bank loans, which requires a measure of the risk premium based on confidential supervisory data ([JLS13]).

puted at the bank level, without relying on precise definitions of the geographic product markets [ADZ14]. For the international-oriented banks in our sample, it is particularly difficult to define geographic markets as they often operate in number of jurisdictions and product markets. Second, the Lerner index measures a bank's pricing power or mark-up and better captures the theoretical concept of bank franchise value ([BDS13]). In addition, the Lerner index utilizes information on both the asset and liability sides of bank balance sheets, as it captures both profits (generated with bank assets) and costs of bank operations ([ADZ14]).

Policy Measures and Cyclical Variables

For the empirical analysis, we control for both monetary and fiscal policy measures. Monetary policy is measured by 3-month short-term interest rates (OECD) and central banks' claims on financial institutions (IMF IFS). Fiscal policy is measured by the ratio of government structural balances to potential GDP, to proxy the fiscal stance (IMF WEO). We use GDP growth as a proxy for cyclical conditions in the economy.

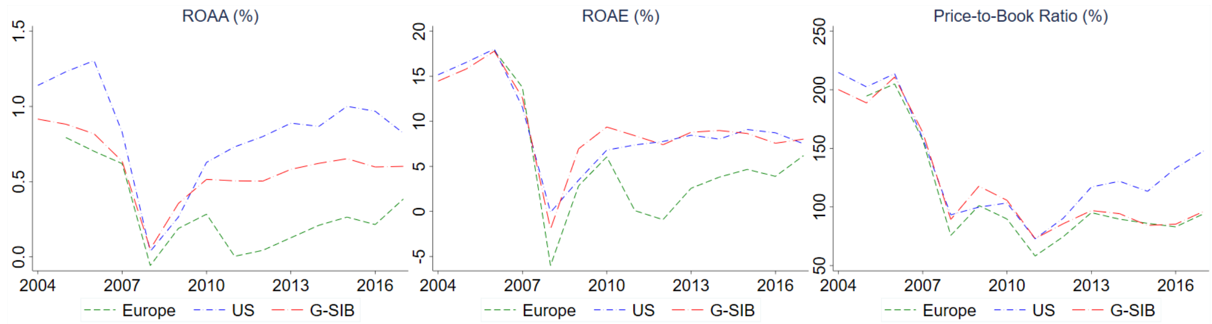
3.3.2 Stylized Facts

Bank profitability, measured by ROAA, ROAE, and price-to-book ratio, all declined sharply during the 2007-2009 GFC (Figure 3.3). In general, U.S. banks have recovered faster than European banks post-crisis partly because European banks experienced another sharp decline in book profitability and price-to-book ratios during the 2012-2014 European Sovereign Debt Crisis. Interestingly, while U.S. banks' book return (ROAA and ROAE) were not affected by the European Sovereign Debt Crisis, their price-to-book ratio experienced a sizable drop. This suggests that while the actual impact of the European crisis on book profitability was regional, it influenced investor's perception of banks' capacity to generate future profits globally. It should be noted that none of the profitability measures have returned to pre-crisis levels[30].

For systemic and idiosyncratic risk measures, we also observe a clear impact from the GFC and the European Sovereign Debt Crisis, where risks became elevated (Figure

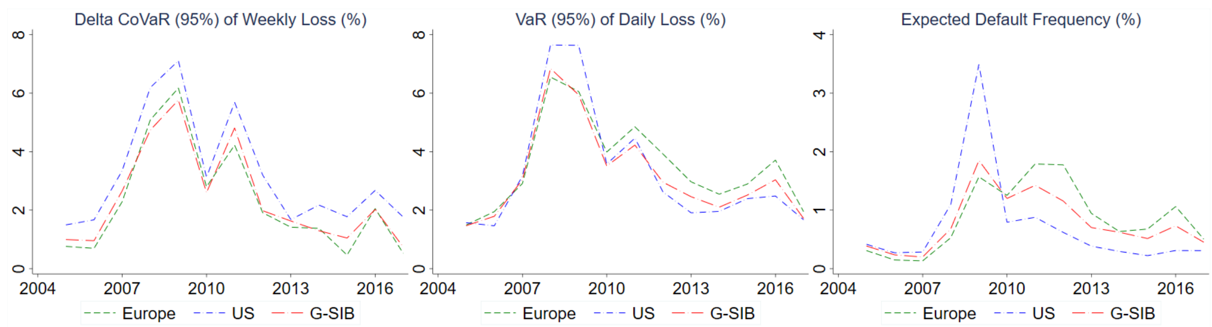
[30]Risk-adjusted ROAA and ROAE display similar dynamics.

Figure 3.3: Bank Profitability and Price-to-Book Ratio



Notes: See Appendix 5.3 for sources and definitions of the variables. The figures are weighted using asset share for each group of banks. The green, blue, and red lines denote European banks, U.S. banks, and Global Systemically Important Banks (GSIBs), respectively. Sources: S&P Global Market Intelligence's SNL database and IMF staff calculations.

Figure 3.4: Systemic and Idiosyncratic Risks



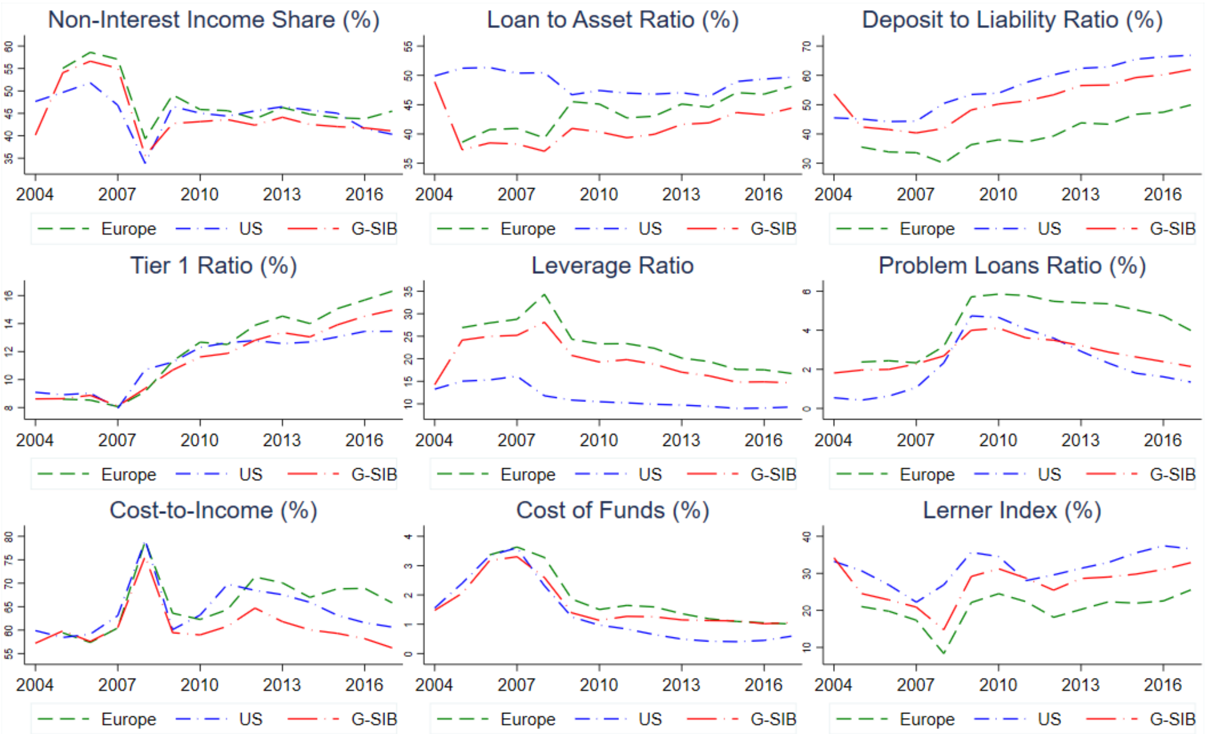
Notes: The figures are weighted using asset share for each group of banks. The green, blue, and red lines denote European banks, U.S. banks, and Global Systemically Important Banks (GSIBs), respectively. Sources: Bloomberg, Moody's Analytics, and IMF staff calculations.

3.4). For systemic risk measured by the $\Delta CoVaR$, U.S. banks tend to have a higher contribution to systemic risk, compared with European banks[31]. For idiosyncratic risks measured by the VaR and the EDF, they appeared elevated for U.S. banks during the 2007-2009 GFC, but were overtaken by European banks during the 2012-2014 European Sovereign Debt Crisis.

On bank characteristics, U.S. banks tend to have higher LTA and deposit-to-liability ratios during the 2004-2017 sample period (Figure 3.5). As expected, GSIBs have the lowest LTA ratios, as a high proportion of their balance sheets are devoted to investment

[31]It should be noted that both $\Delta CoVaR$ measures (with regional and global state variables) show that U.S. banks tend to have a higher contribution to systemic risks, in part because equity returns in the U.S. are more correlated with global equity returns.

Figure 3.5: Bank Business Models and Characteristics



Notes: See Appendix 5.3 for sources and definitions of the variables. The figures are weighted using asset share for each group of banks. The green, blue, and red lines denote European banks, U.S. banks, and Global Systemically Important Banks (GSIBs), respectively. The leverage ratio is constructed as the ratio of asset to equity. Sources: S&P Global Market Intelligence's SNL database and IMF staff calculations.

banking and other non-traditional banking business[32]. In general, LTA ratios have increased for banks since the crisis. While the NII share has declined since the crisis, it has stabilized more recently. European banks appear to have higher leverage and higher problem loan ratios compared with U.S. banks and GSIBs, despite a decline since the peak of the crisis. The Tier 1 ratio has risen markedly for banks since the crisis, in part due to tighter regulations. More recently, cost efficiency, measured by the cost-to-income ratio, has improved for GSIBs and U.S. banks. Similarly, funding costs have declined for banks, with U.S. banks enjoying the lowest cost of funds on average. Finally, the Lerner index suggests that U.S. banks have higher pricing power compared with European banks and GSIBs.

[32] Average LTA ratios for European, U.S., and Asian GSIBs are 43%, 32%, and 48%, respectively, in 2017. Similarly, average NII ratios for European, U.S., and Asian GSIBs are 48%, 58% and 39%, respectively, suggesting that U.S. GSIBs are more involved in investment banking and other non-traditional banking business.

3.3.3 Hypotheses and Empirical Methodology

Based on our stylized theoretical model, we derive four testable hypotheses to be examined empirically. The relationship between idiosyncratic risks and bank profitability and that between profitability and its determinants are directly based on the five propositions from the theoretical model. Furthermore, we are interested in understanding if the predicted relationship between bank profitability and idiosyncratic risks could be extended to systemic risks empirically.

Hypothesis 1. *Low profitability is associated with high idiosyncratic risks (Proposition 3 & 5) and high contribution to systemic risks.*

Hypothesis 2. *High NII share is associated with high idiosyncratic risks (Proposition 4) and high contribution to systemic risks for less retail-oriented banks*[33].

Hypothesis 3. *High leverage is associated with high idiosyncratic risks (Proposition 6) and high contribution to systemic risks.*

The hypotheses relate to bank business models and the impact of the source of bank profitability on risks. In addition, we are also interested in examining the effect of the Lerner index (market power) and the reliance on wholesale funding on idiosyncratic and systemic risks.

The main empirical approach to examine the three hypotheses on bank profitability, financial stability, and business models was a panel regression setup that controls for business models, bank characteristics, as well as policy variables and cyclical conditions in the economy. It was estimated with the Arellano-Bover/Blundell-Bond linear dynamic panel-data estimator with robust standard errors[34], specified as follows:

$$Y_{kjt} = \delta Y_{kj,t-1} + \vartheta_k + \phi' X_{kjt} + \Gamma' M_{jt} + \varepsilon_{kjt} \tag{3.13}$$

[33] *See the discussion of systemic risks in Section 3.2*

[34] A dynamic panel regression is specified due to the persistence in systemic and idiosyncratic risks. The Arellano-Bover/Blundell-Bond system estimator is an extension of the Arellano-Bond estimator that accommodates large autoregressive parameters and a large ratio of the variance of the panel-level effect to the variance of idiosyncratic error. The Arellano-Bover/Blundell-Bond system estimator is designed for datasets with many panels and few periods, which is the case for our dataset. An alternative fixed-effect static panel was specified as a robustness check and the results were found to be broadly similar.

where Y_{kjt} captures either risks or profitability for bank k, headquartered in country j at time t. To take into account bank-specific conditions, we include a set of bank-fixed effects (ϑ_k) and a vector of (time-varying) bank-specific indicators X_{kjt}.

In the panel regression estimations, we examine the determinants of financial stability or risks, measured by the VaR, the EDF, and the $\Delta CoVaR$ (Y_{kjt}). Key bank-specific variables (X_{kjt}) include bank profitability, the share of NII, and the interaction term between NII and the LTA ratio. We also consider the bank-specific problem loan ratio, leverage, and the Lerner index. In addition, we control for policy and cyclical variables (M_{jt}), capturing monetary policy (short-term interest rates), fiscal policy (government structural balance), and GDP growth[35].

3.4 Empirical Findings

This section quantifies the impact of profitability on financial stability using bank-level data for publicly traded U.S., European banks, and Global Systemically Important Banks (GSIBs). As mentioned earlier, financial stability is measured by both idiosyncratic and systemic risks. Idiosyncratic risk is captured by the VaR (5 percent) and Moody's EDF, and systemic risk is captured by the $\Delta CoVaR$.

3.4.1 Idiosyncratic risk

As predicted by our stylized theoretical model in Section II, empirical results over the 2004-2017 sample period reveal that profitability (ROAA) and the price-to-book ratio (charter value) are negatively associated with banks' idiosyncratic risk, measured by VaR (Table 3.1), confirming hypothesis 1. As banks' book profitability or charter value improves, they have more "skin in the game" and they are less willing to engage in risk-taking behavior. This finding suggests that, on average, the charter value channel dominates as higher profits and consequently higher capital is associated with less risk-taking by banks. The effect of profitability on VaR is also economically significant. A one

[35] We do not divide our sample into pre-crisis (2004-2007) and post-crisis (2008-2017) sub-periods due to data limitation. The dynamic panel regression requires the third or deeper lags as instruments, which leaves us with only 2007's observations at best for a pre-crisis analysis.

standard deviation increase in ROAA, for instance, is associated with a 0.64 percentage point decrease in VaR, or about a quarter of the median VaR (2.57 percent). Similarly, a one standard deviation rise in the price-to-book ratio is associated with a 0.59 percentage point decline in VaR, which is also sizable.

In general, a higher NII share is significantly associated with higher VaR. However, the interaction term between the share of NII and the LTA ratio suggests that a higher NII share in retail-oriented banks tends to be associated with a decline in VaR. This empirical finding is consistent with our theoretical prediction that the negative impact from non-interest income is dependent on the value of the LTA ratio, confirming hypothesis 2. One explanation based on our theoretical model is that retail-oriented banks tend to engage in more traditional or fee-based NII activities due to their existing retail client base, instead of more-risky market-based NII activities. Therefore, the marginal effect of NII on bank risks is dependent on the retail orientation of banks, and there could be some diversification benefits for retail-oriented banks to move toward traditional NII activities. Our empirical findings, based on a cross-country sample, are consistent with the earlier work on the German banking sector by [Koh14], which suggested that diversification into NII activities were more beneficial for retail-oriented banks such as savings and cooperative banks.

Higher market power, as measured by the Lerner index, is associated with lower VaR. One potential explanation is that profitable banks have higher charter value and are therefore less willing to engage in risk-taking behavior (e.g., [Kee90, BKT09]). Higher leverage or lower equity-to-asset ratios are associated with higher VaR, confirming hypothesis 3. Furthermore, a high reliance on wholesale funding (low deposit-to-liability ratios) and elevated problem loan ratios tend to be associated with higher idiosyncratic risk as measured by VaR. Finally, a favorable macroeconomic environment, as measured by high GDP growth, is often associated with lower VaR[36].

The empirical findings based on the other measure of bank-specific idiosyncratic risk,

[36] On monetary policy, the empirical result suggests that monetary easing measured by lower short-term interest rate in our sample period from 2004 to 2017 is associated with lower bank-specific idiosyncratic risks, which suggests that the post-crisis monetary policy response by central banks were effective. However, this finding does not bear conclusion on future paths of monetary policy or how central banks should set monetary policy going forward.

the EDF, is similar. As predicted by the theoretical model, both the ROAA and the price-to-book ratios are negatively associated with the one-year ahead EDF (Table 3.2), confirming hypothesis 1.

While a high NII ratio is generally associated with a high default probability, there appears to be diversification benefits for retail-oriented banks, confirming hypothesis 2. As mentioned earlier, one explanation offered by our theoretical model is that retail-oriented banks tend to engage in fee-based NII activities (less risky) due to their existing retail client base, instead of market-based NII activities such as investment banking and securitization (riskier). Our empirical finding on the relationship between default probability and diversification is consistent with earlier work on the U.S. banking sector by [DT13].

Similarly, as was the case for VaR, a higher problem loan ratio is associated with higher default probability for banks. There is some evidence that a higher leverage ratio is significantly associated with higher EDF, confirming hypothesis 3. Finally, a favorable macroeconomic environment is associated with declining bank default probabilities.

3.4.2 Systemic Risks

In addition to bank idiosyncratic risks, we also consider the relationship between bank profitability and systemic risks. Empirical results reveal that profitability (ROAA) and the price-to-book ratios (charter value) are negatively associated with banks' contribution to systemic risk ($\Delta CoVaR$; see Table 3.3). As firm's current book profitability or charter value improves, their contribution to systemic risk tend to decline (hypothesis 1). One intuitive explanation for this finding is that, as banks' charter value increases, they engage in less risk-taking at the individual bank level, and thereby reducing their systemic risk contribution. While some analysis that examined crisis episodes suggests that high profits in good times could be an indicator of systemic tail risk in bad times ([MNP18]), our results suggest that on average the charter value channel dominates as higher profits and consequently higher capital is associated with less risk-taking by banks when we consider a full sample that embeds both crisis and normal times from 2004 to 2017.

Like the findings for idiosyncratic risks, a higher share of NII is generally associated

Table 3.1: Empirical Results: Determinants of Idiosyncratic Risk (VaR)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	VaR	VaR	VaR	VaR	VaR	VaR
ROAA (%)	-0.841***		-1.024***		-1.036***	
	(0.150)		(0.153)		(0.161)	
Price-to-Book Ratio (%)		-0.00917***		-0.0117***		-0.0112***
		(0.00115)		(0.00127)		(0.00114)
Non-Interest Income Share (%)	0.0120*	0.0127*	0.00693	0.00309	0.0215**	-0.00353
	(0.00709)	(0.00720)	(0.00868)	(0.00952)	(0.00961)	(0.00823)
NII Share (%) X Loan-to-Asset Ratio (%)	-0.000227*	-0.000344***	-0.000253*	-0.000138	-0.000421***	-4.18e-05
	(0.000130)	(0.000130)	(0.000148)	(0.000156)	(0.000144)	(0.000152)
Tier 1 Ratio (%)	-0.0598***	-0.107***	-0.0417**	-0.104***	-0.0550***	-0.0900***
	(0.0203)	(0.0201)	(0.0195)	(0.0213)	(0.0166)	(0.0189)
Problem Loans Ratio (%)	0.0434**	0.0479**	0.0460**	0.0514***	0.0542***	0.0518***
	(0.0211)	(0.0215)	(0.0180)	(0.0172)	(0.0191)	(0.0187)
Real GDP growth rate (%)	-0.229***	-0.275***	-0.226***	-0.280***	-0.235***	-0.284***
	(0.0267)	(0.0302)	(0.0264)	(0.0291)	(0.0269)	(0.0286)
ST interest rate (%)	0.116***	0.108***	0.180***	0.222***	0.186***	0.226***
	(0.0300)	(0.0296)	(0.0184)	(0.0190)	(0.0175)	(0.0164)
Gov Structural Balance/Potential GDP	-0.107***	-0.0835***	-0.124***	-0.0869***	-0.114***	-0.0735***
	(0.0179)	(0.0192)	(0.0190)	(0.0195)	(0.0188)	(0.0179)
Lerner Index (%)	-0.0302***	-0.0444***				
	(0.00925)	(0.00908)				
Deposit-to-Liability Ratio (%)			-0.0122**	-0.00983*		
			(0.00479)	(0.00558)		
Leverage Ratio					0.0120	0.0317***
					(0.00912)	(0.0119)
Weekly Delta CoVaR = L	0.266***	0.267***	0.272***	0.331***	0.268***	0.351***
	(0.0371)	(0.0398)	(0.0345)	(0.0372)	(0.0411)	(0.0345)
Observations	3,867	3,833	3,919	3,886	3,922	3,889
Lags	3 - 7	3 - 5	3 - 7	3 - 6	3 - 7	3 - 6
No. of Intruments	390	352	390	342	390	391
No. of Banks	387	386	389	388	389	388
Hansen p-Value	0.425	0.206	0.414	0.113	0.416	0.482
AR2 p-Value	0.000775	7.82e-07	0.00502	2.63e-06	0.00820	3.06e-06
AR3 p-Value	0.00417	2.09e-06	0.00935	3.47e-08	0.0112	7.78e-09

[a] *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. See Appendix 5.3 for sources and definitions of the variables.

Table 3.2: Empirical Results: Determinants of Idiosyncratic Risk (EDF)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	trEDF	trEDF	trEDF	trEDF	trEDF	trEDF
ROAA (%)	-0.218***		-0.231***		-0.207***	
	(0.0650)		(0.0615)		(0.0559)	
Price-to-Book Ratio (%)		-0.000882**		-0.000661*		-0.000206
		(0.000416)		(0.000341)		(0.000350)
Non-Interest Income Share (%)	0.00647***	0.0100***	0.00506*	0.00666**	0.00820***	0.00395
	(0.00233)	(0.00271)	(0.00281)	(0.00332)	(0.00257)	(0.00332)
NII Share (%) X Loan-to-Asset Ratio (%)	-0.000127***	-0.000187***	-0.000130**	-0.000172***	-0.000114**	-0.000113**
	(4.56e-05)	(4.53e-05)	(5.10e-05)	(5.45e-05)	(5.70e-05)	(5.56e-05)
Tier 1 Ratio (%)	0.000793	-0.0146*	0.00610	-0.00826	-0.00121	-0.00367
	(0.00712)	(0.00748)	(0.00752)	(0.00822)	(0.00683)	(0.00737)
Problem Loans Ratio (%)	0.0303***	0.0370***	0.0294***	0.0338***	0.0279***	0.0366***
	(0.00493)	(0.00570)	(0.00499)	(0.00567)	(0.00445)	(0.00546)
Real GDP growth rate (%)	-0.0623***	-0.0817***	-0.0615***	-0.0812***	-0.0608***	-0.0816***
	(0.0115)	(0.0115)	(0.0114)	(0.0115)	(0.0116)	(0.0113)
ST interest rate (%)	0.0438***	0.0434***	0.0443***	0.0355***	0.0414***	0.0317***
	(0.0110)	(0.0103)	(0.00725)	(0.00672)	(0.00669)	(0.00650)
Gov Structural Balance/Potential GDP	-0.0513***	-0.0469***	-0.0513***	-0.0490***	-0.0520***	-0.0524***
	(0.00719)	(0.00783)	(0.00765)	(0.00861)	(0.00719)	(0.00842)
Lerner Index (%)	0.00143	0.00341				
	(0.00302)	(0.00290)				
Deposit-to-Liability Ratio (%)			-0.000795	-0.00226		
			(0.00160)	(0.00177)		
Leverage Ratio					-0.000684	0.0144***
					(0.00180)	(0.00397)
Weekly Delta CoVaR = L	0.578***	0.623***	0.583***	0.628***	0.601***	0.625***
	(0.0317)	(0.0333)	(0.0314)	(0.0327)	(0.0303)	(0.0304)
Observations	2,971	2,943	3,013	2,987	3,016	2,990
Lags	3 - 5	3 - 5	3 - 5	3 - 5	3 - 5	3 - 5
No. of Intruments	287	287	287	270	287	287
No. of Banks	290	287	292	289	292	289
Hansen p-Value	0.401	0.417	0.367	0.203	0.351	0.370
AR2 p-Value	0.0397	1.14e-07	0.0770	1.62e-07	0.0150	1.11e-07
AR3 p-Value	0.103	8.06e-05	0.111	0.000125	0.0551	0.000130

[a] *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. See Appendix 5.3 for sources and definitions of the variables.

Table 3.3: Empirical Results: Determinants of the Contribution to Systemic Risk ($\Delta CoVaR$)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	covar	covar	covar	covar	covar	covar
ROAA (%)	-0.370***		-0.220***		-0.263***	
	(0.107)		(0.0790)		(0.0805)	
Price-to-Book Ratio (%)		-0.00502***		-0.00292***		-0.00348***
		(0.00101)		(0.000845)		(0.00117)
Non-Interest Income Share (%)	0.0432***	0.0446***	0.00884	0.0172*	0.0420***	0.0409***
	(0.00899)	(0.00882)	(0.00946)	(0.00899)	(0.00863)	(0.00894)
NII Share (%) X Loan-to-Asset Ratio (%)	-0.000659***	-0.000617***	-0.000412**	-0.000495***	-0.000706***	-0.000716***
	(0.000166)	(0.000177)	(0.000181)	(0.000187)	(0.000176)	(0.000184)
Tier 1 Ratio (%)	-0.0363	-0.0677***	-0.0324*	-0.0544***	-0.0562***	-0.0836***
	(0.0221)	(0.0188)	(0.0171)	(0.0195)	(0.0194)	(0.0242)
Problem Loans Ratio (%)	-0.0267***	-0.0183*	-0.0467***	-0.0355***	-0.0261***	-0.0187**
	(0.00863)	(0.0101)	(0.00989)	(0.0108)	(0.00843)	(0.00836)
Real GDP growth rate (%)	-0.237***	-0.260***	-0.213***	-0.241***	-0.230***	-0.244***
	(0.0298)	(0.0295)	(0.0286)	(0.0285)	(0.0283)	(0.0287)
ST interest rate (%)	0.0945***	0.117***	0.0121	0.0316	0.0491***	0.0545***
	(0.0294)	(0.0257)	(0.0204)	(0.0207)	(0.0180)	(0.0200)
Gov Structural Balance/Potential GDP	-0.138***	-0.132***	-0.171***	-0.152***	-0.133***	-0.131***
	(0.0201)	(0.0206)	(0.0209)	(0.0210)	(0.0188)	(0.0222)
Lerner Index (%)	0.0156*	0.0186***				
	(0.00807)	(0.00674)				
Deposit-to-Liability Ratio (%)			-0.0328***	-0.0249***		
			(0.00636)	(0.00589)		
Leverage Ratio					0.00945*	0.0279*
					(0.00537)	(0.0166)
Weekly Delta CoVaR = L	0.134***	0.140***	0.127***	0.138***	0.146***	0.154***
	(0.0216)	(0.0228)	(0.0219)	(0.0220)	(0.0210)	(0.0336)
Observations	3,864	3,830	3,916	3,883	3,919	3,886
Lags	2 - 5	2 - 5	2 - 5	2 - 5	2 - 5	3 - 6
No. of Intruments	371	371	371	371	371	342
No. of Banks	386	385	388	387	388	387
Hansen p-Value	0.318	0.370	0.291	0.318	0.336	0.109
AR2 p-Value	0.739	0.354	0.394	0.234	0.589	0.335

[a.] *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. See Appendix 5.3 for sources and definitions of the variables.

with a higher contribution to systemic risks, and there appears to be diversification benefits for retail-oriented banks that move into NII activities (hypothesis 2). There is also evidence that high leverage (hypothesis 3) and an over-reliance on wholesale funding are associated with higher systemic risks.

It is interesting to note that the Lerner index is positively associated with the contribution to systemic risk, but negatively associated with idiosyncratic risks[37]. One potential explanation is that while higher mark-up is beneficial for banks at the bank-specific level, it could increase risks at the system level due to the excessive market power of some banks. This is consistent with the findings for U.S. banks in [ADZ14], which suggest that higher competition (lower mark-up) reduces systemic risks[38].

These empirical findings suggest that the source and the sustainability of bank profitability could carry important financial stability implications. An over-reliance on leverage, wholesale funding sources, and market-based NII tends to be associated with higher idiosyncratic risk and contribution to systemic risks. Consequently, policy makers and financial stability authorities should pay more attention to the source of bank profitability in assessing the resilience of banks to systemic stress. This should also feed into the design and the calibration of macro-prudential stress tests[39]. Furthermore, the differentiating impact of competition on idiosyncratic and systemic risks calls for policy makers to strike a balance between cost reductions (through bank consolidation) and a compet-

[37]The effect of Lerner index is not driven by bank size. In our sample, the correlation between the Lerner index and market power are weak (at 0.10). In general, there is mixed evidence between bank size and bank market power, both theoretically and empirically. Some studies predict a positive relation (e.g., consider a Cournot competition model). Others predict a negative oneas larger banks tend to operate on a larger (national or international) market whereas smaller ones locally (e.g., smaller banks focus on relationship banking, where they exploit higher rents from asymmetric information and higher switching costs of borrowers). See [BSF06].

[38]The NPL ratio is found to be positively associated with idiosyncratic risk but negatively associated with systemic risk. One potential explanation here is that systemic risk is related to the degree of interconnectedness of one bank with the rest of the market ([DY14, MX17]), and therefore market-based activities are more influential in banks' contribution to systemic risk, compared with retail-based activities that typically determine banks' problem loan ratios. After controlling for common cyclical conditions (GDP growth), the problem loan ratio tends to reflect bank-specific risk appetite and risk management practices, which may be different from that of other banks. As a result, the problem loan ratio could have a low or negative beta compared with general market movements and could be negatively associated with banks' contribution to systemic risks.

[39]Typically, in a stress testing exercise, bank profitability matters through retained earnings and capital adequacy. However, more attention could be paid to the source and the sustainability in a systemic risk analysis of the financial system (for example, by including a detailed profitability analysis alongside the stress testing exercise).

itive and stable banking environment. One approach to facilitate a competitive banking environment is to allow for the entry of new firms instead of raising domestic and foreign entry barriers into the financial sector to unnecessarily high levels ([ADZ14]).

A number of further robustness checks were carried out, including using various lag-specifications as instruments in the dynamic panel regression, re-running the regressions using static panels with fixed effects against lagged independent variables, and including time dummies (year-fixed effects). The results are found to be robust (see Appendix 5.4 for details).

3.5 Policy Implications and Conclusions

This paper investigates the relationship between bank profitability and financial stability, accounting for bank business models and different NII activities. It also examines the importance of the various determinants of banking risks and profitability. The paper first develops a stylized theoretical model that captures bank risks and retail-based and market-based NII activities. It then estimates a panel regression model for 431 publicly traded banks from 2004 to 2017.

The stylized theoretical model establishes the analytical relationship between financial stability and bank profitability, and between financial stability and business models captured by NII activities. The model predicts that idiosyncratic risks, captured by the VaR of equity prices and the EDF, are negatively related to both ROAA and long-term expected profitability (i.e., charter value). Profits reduce risks by providing equity buffers, and by encouraging prudence and reduced risk-taking. In addition, idiosyncratic risk rises with the share of NII activities when the LTA ratio is below a certain threshold. Idiosyncratic risk also increases with the leverage ratio of banks. The theoretical model also predicts that profitability decreases as the problem loan ratio, operating costs, and funding costs increase.

The empirical results confirm the theoretical predictions on bank profitability and financial stability. First, profitability (ROAA) and the price-to-book ratio are negatively associated with both contribution to systemic risk ($\Delta CoVaR$) and idiosyncratic risk measured by VaR and the EDF of banks. Second, a high NII share tends to be associated with higher idiosyncratic risk and contribution to systemic risk when the LTA ratio is low (i.e., when a bank's business model is less retail-oriented), as predicted by the theoretical model. Third, lower competition (high mark-up) is associated with lower idiosyncratic risk but higher contribution to systemic risk. Fourth, the empirical results suggest that high leverage and over-reliance on wholesale funding are associated with higher risks.

These findings raise several interesting issues for policy makers and financial stability authorities. First, the results highlight the need for a sharper distinction between different types of NII activities. In general, market-based NII activities are riskier than retail-based NII activities. This is an important consideration. In a low interest rate environment, banks tend to diversify into NII activities, but this causes a shift in a bank's risk profile. Second, it would be important to account for the impact of bank consolidation on competition and systemic risks. Low competition is associated with high contribution to systemic risk but low idiosyncratic risk. After the recent GFC experience, there was a rise in mergers and acquisitions between banks. While beneficial for banks at the firm level, lower competition as measured by a higher Lerner index appears to be negatively associated with banks' contribution to systemic risk. From a financial stability policy viewpoint, the right balance between cost efficiency and a competitive and stable banking environment is an important consideration. Third, these results highlight the need to evaluate the sustainability of bank profitability. An over-reliance on leverage and wholesale funding are associated with higher idiosyncratic and contribution to systemic risks and thereby lower financial stability. Policy makers and financial stability authorities should pay more attention to the source and the sustainability of bank profitability in the design and the calibration of macro-prudential stress tests and systemic risk analysis. These findings also underscore the importance of the effective and timely implementation of the Basel III framework, the need for well calibrated macro-prudential tools, and to ensure that banks' reliance on wholesale funding and leverage remains prudentially manageable.

CHAPTER 4

Appendix I

4.1 Data Development and Selection

4.1.1 Compustat Firm Data Selection

Table 4.1: Sample Selection for Compustat Data

	Selection	SIC	Obs	Firms
	Compustat		378,983	32,467
–	Financials	6000-6999	287,321	23,888
–	Argricultural	0-999	285,954	23,761
–	Public Service	9000-9999	280,291	23,225
–	R&D=0 for its entire operation periods		148,290	11,395

4.1.2 PatentsView Patent Data Selection

Classification of assignee: 2 - US Company or Corporation, 3 - Foreign Company or Corporation, 4 - US Individual, 5 - Foreign Individual, 6 - US Government, 7 - Foreign Government, 8 - Country Government, 9 - State Government (US). Note: "A" or "1" appearing before any of these codes signifies part interest.

Table 4.2: Sample Selection for Unmatched Patent Data

Selection	Criteria	Obs	Selection	Criteria	Obs
Patent		6,408,293	Application		6,422,963
Country	=US	6,408,291	Series code	$\neq$29	5,993,792
Type	=Utility	5,807,528			
Kind	=A,B1,B2	5,807,519			
Selection	Criteria	Obs	Selection	Criteria	Obs
Assignee		379,565	Merged Patents Files		6,162,134
Type	=2	167,986	Assignee	Missing Info	5,354,027
Organization	Missing name	167,741	Type	=2	2,657,303
			Citation data	Matched	2,560,204
			Application year	$<$ 1975 or $>$ 2017	2,536,585

4.1.3 Firm-Patent Name Matching

Table 4.3: Name Matching Process

Step	Criteria	Fuzzy Match	Obs	Cumulative Matches	File
0			11,398		getCompName.do
1	standard_name	N		5,278	
2	stem_name	N		6,635	
3	stem_name	Y	992	7,304	match_ass.py
4	merged with cleaned patent data			6,705	clean_patent.do

Note that in Step 3, the number of matches with ratio $\geq$ 95 is 468, which I keep with certainty. Matches with ratio $<$ 95 but $\geq$ 93, I check one by one manually. I only kept 201 of them (524).

By matching application year-firm (gvkey) pairs and corresponding fiscal year-firm (gvkey) pairs data from Compustat, the total number of matched patents is 810,451 (unmatched 1,749,753), of firms 4,630 and of firm-years 70,004.

Table 4.4: Selection for Matched PatentsView-Compustat Data

Criteria		Obs	Firms	Patents
total	firm-patent-year obs	929,003	11,395	862,038
total	firm-year obs	147,850	11,395	862,038
Patent=0 for its entire operation periods		70,004	4630	862,038
1997-2016 data		35,691	3,482	604,193
Continuous innovating from 97-16		31,895	3,187	569,304

4.2 Additional Statistics

Total number of observations is 1,040.

Table 4.5: Summary Statistics (Fiscal Year 2015)

	Mean	Median	St. Dev.
Employee	10,071	660	32,508
Revenue (mil \$)	4,320	200	18,113
R&D expense (mil \$)	253	28	1,019
R&D Exp/Net sales	9.74%	13.20%	79.56%
Pre-tax income/Net sales	-20.98	0.00	232.48
Growth rate of net sales	3.93%	2.77%	62.50%
Patents applied	13.36	1.00	130.03
Self-citation ratio	14.28%	8.67%	18.17%

We can see firms get bigger, especially after 2011 and for the top half of firms.

Table 4.6: Percentiles of Size Distribution

Percentile	1997	2002	2007	2011	2016
10	44	40	38	27	26
20	87	93	91	80	73
30	143	152	165	165	131
40	235	250	290	346	290
50	372	431	496	614	630
60	716	793	951	1300	1475
70	1444	1577	2000	2735	3085
80	3360	3522	4700	5672	6966
90	9100	8600	13355	16000	17678

Figure 4.1: Innovation Capacity by Firm Size

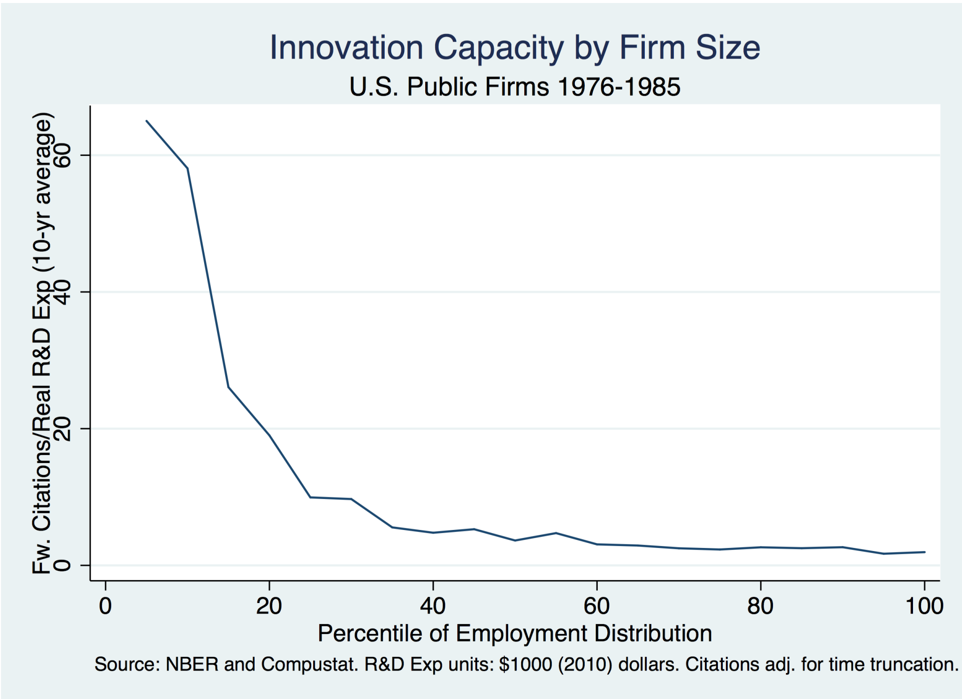
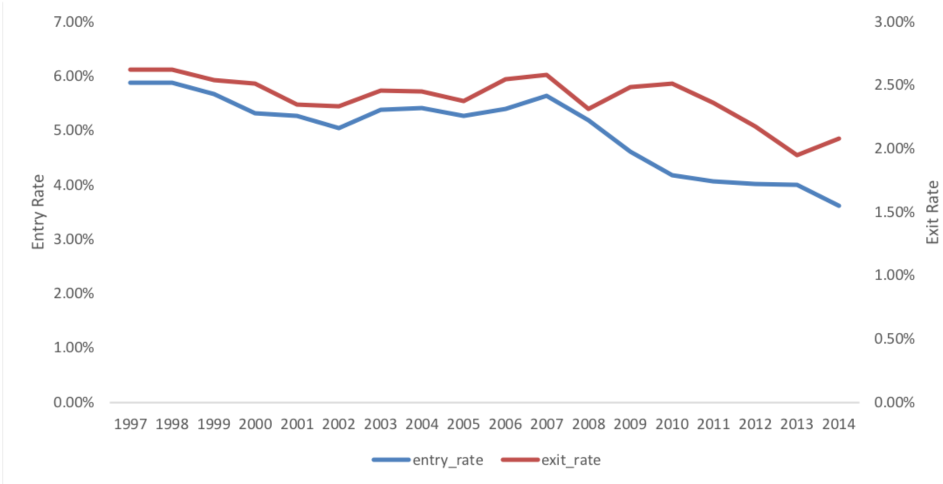


Figure 4.2: Firm Entry and Exit Rates by Year



4.3 Equilibrium Prices and Profits

4.3.1 Final Goods Production

Omitting t

$$\max_{\{y_j\}} A\left(\int_0^1 y_j(t)^{\frac{\epsilon-1}{\epsilon}}dj\right)^{\frac{\epsilon}{\epsilon-1}} - \int_0^1 p_j y_j dj$$

F.O.C. w.r.t. y_j

$$p_j = A^{1-1/\epsilon}\left(\frac{Y}{y_j}\right)^{1/\epsilon}.$$

4.3.2 Intermediate goods production

$$\Pi(n,z) = \max_{\{y_j\}_{j\in\mathbf{J}}} \sum_{j\in\mathbf{J}}[p_j y_j - \frac{w}{z}y_j] = \max_{\{y_j\}_{j\in\mathbf{J}}} \sum_{j\in\mathbf{J}}[A^{1-1/\epsilon}Y^{1/\epsilon}y_j^{1-1/\epsilon} - \frac{w}{z}y_j]$$

Therefore

$$y_j = \left[\frac{(\epsilon-1)z}{\epsilon w}\right]^{\epsilon} YA^{\epsilon-1}. \tag{4.1}$$

By substituting (4.1) into final good production (2.6), we have

$$w = A\frac{\epsilon-1}{\epsilon}\bar{z} \tag{4.2}$$

where $\bar{z} \equiv \left[\int_0^M \sum_{j\in\mathbf{J}_f} z_f^{\epsilon-1} d_f\right]^{\frac{1}{\epsilon-1}} = \left[\int_0^1 z_j^{\epsilon-1} d_j\right]^{\frac{1}{\epsilon-1}}$.

Using (4.2) to rewrite y_j in terms of $\hat{z} = z/\bar{z}$, we have

$$y_j = \hat{z}^{\epsilon} A^{-1} Y \text{ and } p_j = A\hat{z}^{-1} \text{ and } l_j = y_j/z = \hat{z}^{\epsilon-1} A^{-1} Y/\bar{z}.$$

Therefore net sales is $p_j y_j = \hat{z}^{\epsilon-1} Y$ and profit $\pi_j = p_j y_j - w l_j = \hat{z}^{\epsilon-1} Y/\epsilon$ and

$$\Pi(n, z) = n\pi_j = nY\hat{z}^{\epsilon-1}/\epsilon.$$

4.3.3 Labor Market Clearing

From labor market clearing $\int_0^1 l_j dj = 1$, we have

$$Y = A\bar{z}.$$

Thus,

$$\Pi(n, z) = n\pi_j = nY\hat{z}^{\epsilon-1}/\epsilon = n\bar{z}A\hat{z}^{\epsilon-1}/\epsilon.$$

The profit share of total output is, using the fact that $\int_0^1 \hat{z}_j^{\epsilon-1} dj = 1$,

$$\frac{\int_0^1 \pi_j dj}{Y} = \frac{Y/\epsilon}{Y} = 1/\epsilon$$

and labor income share is $1 - 1/\epsilon$.

4.3.4 Euler Equation

From the maximization problem of households, we have the standard Euler equation

$$g = \frac{\dot{C}}{C} = \frac{\dot{Y}}{Y} = \frac{\dot{\bar{z}}}{\bar{z}} = \frac{r - \rho}{\theta}.$$

4.4 Lemma and Proof

Lemma 1

Proof. The firm's value function $\mathbf{V}(n, \hat{z}, \bar{z})$ is defined as

$$
\begin{aligned}
r\mathbf{V}(n, \hat{z}, \bar{z}) = \max_{k_I, k_x} & \underbrace{n\bar{z}A\hat{z}^{\epsilon-1}/\epsilon}_{\text{profit}} - \underbrace{n(k_I\hat{z}^{\epsilon-1} + k_x)\bar{z}}_{\text{R\&D cost}} \\
& + \underbrace{F_I(k_I)[\mathbf{V}(n, \hat{z}(1+\lambda), \bar{z}) - \mathbf{V}(n, \hat{z}, \bar{z})]}_{\text{return from internal R\&D}} \\
& + \underbrace{n\tau[\mathbf{V}(n-1, \hat{z}, \bar{z}) - \mathbf{V}(n, \hat{z}, \bar{z})]}_{\text{creative destruction}} \\
& + \underbrace{nF_x(k_x, n)\left[\mathbb{E}_{\hat{z}'}\mathbf{V}(n+1, \frac{n\hat{z} + \hat{z}'(1+\eta)}{n+1}, \bar{z}) - \mathbf{V}(n, \hat{z}, \bar{z})\right]}_{\text{return from external R\&D}} \\
& - \frac{\partial \mathbf{V}}{\partial \hat{z}}(n, \hat{z}, \bar{z})g\hat{z} + \frac{\partial \mathbf{V}}{\partial \bar{z}}(n, \hat{z}, \bar{z})g\bar{z} \\
& + \underbrace{\varphi\left[\iota A\hat{z}^{\epsilon-1}/\epsilon - k_I\hat{z}^{\epsilon-1} - k_x\right]n\bar{z}}_{\text{financial constraint}}
\end{aligned}
$$

where the second- and third-to-last terms use $\dot{\bar{z}} = g\bar{z}$ and $\dot{\hat{z}} = -g\hat{z}$ in balance growth path equilibrium.

We can easily verify that $\mathbf{V}(n, \hat{z}, \bar{z}) = \bar{z}V(n, \hat{z})$, where $V(n, \hat{z})$ is defined in (2.16).

■

Lemma 2

Proof. For a some time interval Δ, the evolution of the aggregated productivity $\bar{z}$ is given by

$$
\begin{aligned}
&\bar{z}^{\epsilon-1}(t+\Delta) = \int_0^\infty z^{\epsilon-1}\tilde{\phi}_{t+\Delta}(z)dz \\
&= \int_0^\infty \left[\Delta\tau(z(1+\eta))^{\epsilon-1} + \Delta\tilde{I}(z)(z(1+\lambda))^{\epsilon-1} + (1 - \Delta\tau - \Delta\tilde{I}(z))z^{\epsilon-1}\right]\tilde{\phi}_t(z)dz
\end{aligned}
$$

where $\tilde{\phi}_t(z)$ is the marginal distribution of z at time t, and $\tilde{I}(z)$ is the internal innovation rate by firms with productivity z.

The differential becomes

$$
\begin{aligned}
&\frac{\bar{z}^{\epsilon-1}(t+\Delta) - \bar{z}^{\epsilon-1}(t)}{\Delta} = \\
&\int_0^\infty \left[\tau((z(1+\eta))^{\epsilon-1} - z^{\epsilon-1}) + \tilde{I}(z)((z(1+\lambda))^{\epsilon-1} - z^{\epsilon-1})\right]\tilde{\phi}_t(z)dz.
\end{aligned}
$$

Normalizing by dividing both sides by $\bar{z}^{\epsilon-1}(t)$, it becomes

$$\frac{\bar{z}^{\epsilon-1}(t+\Delta)-\bar{z}^{\epsilon-1}(t)}{\Delta\bar{z}^{\epsilon-1}(t)} = \int_0^\infty \left[\tau((\hat{z}(1+\eta))^{\epsilon-1}-\hat{z}^{\epsilon-1})+I(\hat{z})((\hat{z}(1+\lambda))^{\epsilon-1}-\hat{z}^{\epsilon-1})\right]\phi_t(\hat{z})d\hat{z}$$

where $\phi(q)=\sum_{n=1}^\infty h(n,q)$ is the unconditional distribution of $\hat{z}$

and $I(\hat{z}) = M\sum_{n=1}^\infty F_I(k_I(n,\hat{z}))h(n,\hat{z})/\phi(\hat{z})$ is the internal innovation rate conditional on being type $\hat{z}$ firms.

Take Δ to 0. We have

$$g = \frac{\overbrace{\tau\mathbb{E}_{\hat{z}}\left[(\hat{z}(1+\eta))^{\epsilon-1}-\hat{z}^{\epsilon-1}\right]}^{\text{agg. external R\&D and entry}}+\overbrace{\mathbb{E}_{\hat{z}}\left\{I(\hat{z})\left[(\hat{z}(1+\lambda))^{\epsilon-1}-\hat{z}^{\epsilon-1}\right]\right\}}^{\text{agg. internal R\&D}}}{\epsilon-1}.$$

Simplify and using $\mathbb{E}_{\hat{z}}(\hat{z}^{\epsilon-1})=1$, we have equation (2.20).

The aggregate creative destruction rate τ is derived from the definition.

■

Lemma 3

Proof. When $k_x>0$, from the FOC of intermediate goods producers' value function

$$\begin{aligned}
n\hat{z}^{\epsilon-1}(1+\varphi) &= \alpha_I\beta_I k_I^{\beta_I-1}[V(n,\hat{z}(1+\lambda))-V(n,\hat{z})]\\
n(1+\varphi) &= \alpha_x\beta_x k_x^{\beta_x-1}n^{\gamma+1}\left[\mathbb{E}_{\hat{z}'}V(n+1,\frac{n\hat{z}+\hat{z}'(1+\eta)}{n+1})-V(n,\hat{z})\right]\\
\iota A\hat{z}^{\epsilon-1}/\epsilon &\geq k_I\hat{z}^{\epsilon-1}+k_x+\Phi \text{ with equality when } \varphi>0.
\end{aligned}$$

Then we have

$$\begin{aligned}
k_I^* &= \left[\frac{\alpha_I\beta_I[V(n,\hat{z}(1+\lambda))-V(n,\hat{z})]}{n\hat{z}^{\epsilon-1}(1+\varphi)}\right]^{\frac{1}{1-\beta_I}}\\
k_x^* &= \left[\frac{\alpha_x\beta_x[\mathbb{E}_{\hat{z}'}V(n+1,\frac{n\hat{z}+\hat{z}'(1+\eta)}{n+1})-V(n,\hat{z})]}{n^{-\gamma}(1+\varphi)}\right]^{\frac{1}{1-\beta_x}}
\end{aligned}$$

■

Lemma 5 (Firm Growth Rate). *Let $g_f \equiv \mathbf{E}\left(\dot{Q}_f/Q_f\right)$ be the average growth rate of a firm with $n\hat{z}^{\epsilon-1} \equiv Q_f$. Then in equilibrium,*

$$g_f = \frac{nx(n,\hat{z})\left\{(n+1)\mathbb{E}\left[\left(\frac{n\hat{z}+(1+\eta)\hat{z}'}{n+1}\right)^{\epsilon-1}\right] - Q_f\right\}}{Q_f} + I(n,\hat{z})\lambda - \tau \tag{4.3}$$

where $x(n,\hat{z}) = F_x(k_x(n,\hat{z}))$ and $I(n,\hat{z}) = F_I(k_I(n,\hat{z}))$.

Proof. Consider a small time interval Δ, then $Q_f = n\hat{z}^{\epsilon-1}$ follows

$$\begin{aligned} Q_f(t+\Delta) =& nx(n,\hat{z})\Delta(n+1)\mathbb{E}\left\{\left[\frac{n\hat{z}+(1+\eta)\hat{z}'}{n+1}\right]^{\epsilon-1}\right\} + I(n,\hat{z})\Delta[Q_f(t)(1+\lambda)] \\ &+ n\tau\Delta[Q_f(t) - \frac{Q_f(t)}{n}] \\ &+ (1 - nx(n,\hat{z})\Delta - I(n,\hat{z})\Delta - n\tau\Delta)Q_f(t) \end{aligned}$$

where $x(n,\hat{z}) = F_x(k_x(n,\hat{z}))$ and $I(n,\hat{z}) = F_I(k_I(n,\hat{z}))$.

Subtract $Q_f(t)$ from both sides of the equation, divide both sides by $\Delta Q_f(t)$, and then take Δ to 0, we have Equation (4.3), using the fact that $g_f = \lim_{\Delta\to 0}\frac{Q_f(t+\Delta)-Q_f(t)}{\Delta Q_f(t)}$.

■

Lemma 6 (Stationary Distribution). *Denote the joint distribution by $H(n,\hat{z}) = Prob(q \leq \hat{z}, \tilde{n} = n)$. Its density $h(n,\hat{z})$ satisfies the Kolmogorov forward equation (for $n \geq 2$)*

$$\begin{aligned} &\dot{h}(n,\hat{z}) = \\ &-\underbrace{\left[I(n,\hat{z})h(n,\hat{z}) - \frac{1}{1+\lambda}I(n,\frac{\hat{z}}{1+\lambda})h(n,\frac{\hat{z}}{1+\lambda})\right]}_{\text{outflow due to internal R\&D}} - \underbrace{nx(n,\hat{z})h(n,\hat{z})}_{n\text{-sized firm to } n+1} - \underbrace{n\tau h(n,\hat{z})}_{n\text{-size firm to } n-1} \\ &+ \underbrace{\int_0^{\frac{n\hat{z}}{1+\eta}}(n-1)x(n-1,\frac{n\hat{z}-q(1+\eta)}{n-1})h(n-1,\frac{n\hat{z}-q(1+\eta)}{n-1})\phi(q)dq}_{\text{inflow due to external R\&D by size } n-1 \text{ firms}} \\ &+ \underbrace{(n+1)\tau h(n+1,\hat{z})}_{\text{creative destruction of size-}n+1\text{ firms}} + \underbrace{\frac{\partial[g\hat{z}h(n,\hat{z})]}{\partial\hat{z}}}_{\text{extensive margin}} \end{aligned} \tag{4.4}$$

where $x(n,q) = F_x(k_X(n,q),n)$, $I(n,q) = F_I(K_I(n,q))$ and $q \sim \phi(q) = \sum_{n=1}^{\infty} h(n,q)$.

One can derive a similar equation for $n = 1$.

The total measure of firms M satisfies

$$1 = M \sum_{n=1}^{\infty} n \int_0^{\infty} h(n, \hat{z}) d\hat{z}.$$

Proof. for $n \geq 2$

Consider banks with $= n$ and $\leq \hat{z}$

Outflow

1. n-sized firms internally innovate to $> \hat{z}$
2. n-sized firms externally innovate to $n + 1$
3. n-sized firms are creatively destructed to $n - 1$.

Inflow

1. $(n-1)$-sized firms externally innovate to $= n$ but $\leq \hat{z}$
2. Firms with $= n + 1, \leq \hat{z}$ are creatively destructed to $= n$
3. Firms with $= \hat{z}$ become $< \hat{z}$ when they fail to innovate as $\bar{z}$ grows.

Let $x(n, q) = F_x(k_X(n, q), n)$ and $I(n, q) = F_I(K_I(n, q))$. Consider a small time interval from t to $t + \Delta$

$$\begin{aligned} H_{t+\Delta}(n, \hat{z}) &= H_t(n, \hat{z}(1 + g\Delta)) \\ &- \underbrace{\int_{\hat{z}(1+g\Delta)/(1+\lambda)}^{(1+g\Delta)\hat{z}} \Delta I(n, q) h(n, q) dq}_{\text{outflow due to internal R\&D}} - \underbrace{\int_0^{\hat{z}(1+g\Delta)} \Delta n x(n, q) h(n, q) dq}_{\text{outflow: n-sized firm external R\&D to } n+1} \\ &- \underbrace{\Delta n \tau H_t(n, \hat{z})}_{\text{out: creative destruction of n-sized firm}} + \underbrace{\Delta X(n - 1, \hat{z})}_{\text{inflow due to external R\&D by size } = n-1 \text{ firms}} \\ &+ \underbrace{(n + 1)\Delta \tau H(n + 1, \hat{z}(1 + g\Delta))}_{\text{creative destruction of size-}n+1\text{ firms}} \end{aligned} \tag{4.5}$$

where

$$X(n-1,\hat{z}) = \int_0^{\hat{z}} \int_0^{\infty} (n-1)x(n-1, \frac{n\tilde{z} - q'(1+\eta)}{n-1})h(n-1, \frac{n\tilde{z} - q'(1+\eta)}{n-1})\phi(q')dq'd\tilde{z}$$

$$q \sim \phi(q) = \sum_{n=1}^{\infty} h(n,q) \text{ the marginal distribution of } \hat{z}.$$

Note that the second-to-last line comes from the convolution formula.

1. After a successful external innovation of a firm of size $= n-1$, its post-innovation productivity should be $\tilde{z} \leq \hat{z}$.

2. Let its original productivity be q and the productivity of the product line it creatively destructs be q'; therefore, we need $n\tilde{z} \equiv (n-1)q + q'(1+\eta) < n\hat{z}$.

3. q' is random draw from the average relative productivity distribution $H(\hat{z})$, so q' is independent of q.

4. The pdf of the new productivity $\tilde{z}$ follows

$$\int_0^{\infty} (n-1)x(n-1, \frac{n\tilde{z} - q'(1+\eta)}{n-1})h(n-1, \frac{n\tilde{z} - q'(1+\eta)}{n-1})\phi(q')dq'.$$

5. Therefore, the total outflow due to external innovation is

$$\int_0^{\hat{z}} \int_0^{\infty} (n-1)x(n-1, \frac{n\tilde{z} - q'(1+\eta)}{n-1})h(n-1, \frac{n\tilde{z} - q'(1+\eta)}{n-1})\phi(q')dq'd\tilde{z}.$$

Subtracting both sides of (4.5) by $H_t(n,\hat{z})$, dividing both sides by Δ, taking Δ to 0, and using the fact that

$$\lim_{\Delta \to 0} \frac{H_{t+\Delta}(n,\hat{z}) - H_t(n,\hat{z})}{\Delta} \equiv \dot{H}(n,\hat{z})$$

and

$$\lim_{\Delta \to 0} \frac{H_t(n,\hat{z}(1+g\Delta)) - H_t(n,\hat{z})}{\Delta} = g\hat{z}h(n,\hat{z})$$

we can derive the flow equation.

$$
\begin{aligned}
&\dot{H}(n,\hat{z}) = \\
&-\int_{\hat{z}/(1+\lambda)}^{\hat{z}} I(n,q)h(n,q)dq - \int_0^{\hat{z}} nx(n,q)h(n,q)dq \\
&-\tau n H(n,\hat{z}) + X(n-1,\hat{z}) + (n+1)\tau H(n+1,\hat{z}) + g\hat{z}h(n,\hat{z}).
\end{aligned}
$$

Now differentiate w.r.t. $\hat{z}$ using the Leibniz integral rule

$$
\begin{aligned}
&\dot{h}(n,\hat{z}) = \\
&-\left[I(n,\hat{z})h(n,\hat{z}) - \frac{1}{1+\lambda}I(n,\frac{\hat{z}}{1+\lambda})h(n,\frac{\hat{z}}{1+\lambda})\right] - n(x(n,\hat{z})+\tau)h(n,\hat{z}) \\
&+ \varrho(n-1,\hat{z}) + (n+1)\tau h(n+1,\hat{z}) + \frac{\partial[g\hat{z}h(n,\hat{z})]}{\partial\hat{z}}
\end{aligned}
$$

$$
\begin{aligned}
\varrho(n-1,\hat{z}) &= \int_0^{\frac{n\hat{z}}{1+\eta}} (n-1)x(n-1,\frac{n\hat{z}-q(1+\eta)}{n-1})h(n-1,\frac{n\hat{z}-q(1+\eta)}{n-1})\phi(q)dq \\
q \sim \phi(q) &= \sum_{n=1}^{\infty} h(n,q) \text{ the marginal distribution of } \hat{z}.
\end{aligned}
$$

For $n = 1$, we have

$$
\begin{aligned}
\underbrace{\dot{H}(1,\hat{z})}_{\text{net flow}} = &-\underbrace{\int_{\hat{z}/(1+\lambda)}^{\hat{z}} I(1,q)h(1,q)dq}_{\text{outflow due to internal R\&D}} - \underbrace{\left[\int_0^{\hat{z}} x(1,q)h(1,q)dq + \tau H(1,\hat{z})\right]}_{\text{external R\&D and creative destruction}} \\
&\underbrace{+x_e\int_0^{\hat{z}/(1+\eta)} \phi(q)dq}_{\text{inflow from entrants}} + \underbrace{2\tau H(2,\hat{z})}_{\text{creative destruction of size 2 firms}} + \underbrace{g\hat{z}h(1,\hat{z})}_{\text{extensive margin}} .
\end{aligned}
$$

So the Kolmogorov forward equation is

$$
\begin{aligned}
\dot{h}(1,\hat{z}) = &-\left[I(1,\hat{z})h(1,\hat{z}) - \frac{1}{1+\lambda}I(1,\frac{\hat{z}}{1+\lambda})h(1,\frac{\hat{z}}{1+\lambda})\right] - x(1,\hat{z})h(1,\hat{z}) - \tau h(1,\hat{z}) \\
&+ x_e\phi(\hat{z}/(1+\eta)) + 2\tau h(2,\hat{z}) + \frac{\partial[g\hat{z}h(1,\hat{z})]}{\partial\hat{z}}.
\end{aligned}
$$

■

Proposition 1

Proof. From Lemma 3, a firm with given $(n, \hat{z})$ will have

$$\frac{k_I}{k_x} \propto (1+\varphi)^{\frac{1}{1-\beta_x}-\frac{1}{1-\beta_I}}$$

$\frac{k_I}{k_x}$ is increasing in φ if and only if $\beta_x > \beta_I$.

■

4.5 Myopic Firm Problem

This section lays out the analysis for a myopic profit maximizing firm to derive closed-form solutions.

For a small time interval Δ, a myopic firm maximizes one-period ahead payoffs

$$\begin{aligned}
\max_{k_I,k_x} \ & \Delta[\Pi(n,\hat{z}) - n(k_I\hat{z}^{\epsilon-1} + k_x)] \\
& + (1-r\Delta)\Delta F_I(k_I)[\Pi(n,\hat{z}(1+\lambda)) - \Pi(n,\hat{z})] \\
& + (1-r\Delta)\Delta n F_x(k_x,n)\left\{\mathbf{E}_{\hat{z}'}\Pi\left[n+1, \frac{n\hat{z}+\hat{z}'(1+\eta)}{n+1}\right] - \Pi(n,\hat{z})\right\} \\
& + (1-r\Delta)\Delta n\tau[\Pi(n-1,\hat{z}) - \Pi(n,\hat{z})] \\
& + (1-r\Delta)\Pi(n,\hat{z})
\end{aligned}$$

s.t.

$$\iota A n\hat{z}^{\epsilon-1}/\epsilon \geq n(k_I\hat{z}^{\epsilon-1} + k_x).$$

Keeping only terms related to k_I and k_x, factor out Δ and take $\Delta \to 0$, we have

$$\begin{aligned}
\max_{k_I,k_x} \ & -n(k_I\hat{z}^{\epsilon-1} + k_x) \\
& + F_I(k_I)[\Pi(n,\hat{z}(1+\lambda)) - \Pi(n,\hat{z})] \\
& + nF_x(k_x,n)\left\{\mathbf{E}_{\hat{z}'}\Pi\left[n+1, \frac{n\hat{z}+\hat{z}'(1+\eta)}{n+1}\right] - \Pi(n,\hat{z})\right\}.
\end{aligned}$$

Using $\Pi(n, \hat{z}) = nA\hat{z}^{\epsilon-1}/\epsilon \equiv n\pi(\hat{z})$ where $\pi(\hat{z}) = A\hat{z}^{\epsilon-1}/\epsilon$, we have

$$\begin{aligned}\max_{k_I, k_x} \; & -(k_I\hat{z}^{\epsilon-1} + k_x + \Phi \cdot 1_{k_x>0}) \\ & + F_I(k_I)[\pi(\hat{z}(1+\lambda)) - \pi(\hat{z})] \\ & + F_x(k_x, n)\left\{(n+1)\mathbf{E}_{\hat{z}'}\pi\left[\frac{n\hat{z} + \hat{z}'(1+\eta)}{n+1}\right] - n\pi(\hat{z})\right\} \\ & + \varphi[\iota\pi(\hat{z}) - (k_I\hat{z}^{\epsilon-1} + k_x].\end{aligned} \tag{4.6}$$

When $\epsilon = 2$, we then have linear profit functions. Furthermore, set $A = 1$. Then the maximization problem (4.6) becomes

$$\max_{k_I, k_x} F_I(k_I)\hat{z}\lambda + F_x(k_x, n)(1+\eta) - (k_I\hat{z} + k_x) + \varphi[\iota A\hat{z}/2 - (k_I\hat{z} + k_x)].$$

Note that now $\bar{\hat{z}} = 1$ when $\epsilon = 2$

$$\max_{k_I, k_x} \alpha_I k_I^{\beta_I}\hat{z}\lambda + \alpha_x k_x^{\beta_x} n^{\gamma}(1+\eta) - (k_I\hat{z} + k_x) + \varphi[\iota\hat{z}/2 - (k_I\hat{z} + k_x)].$$

F.O.Cs for firms with $n \geq 1$

$$k_I^* = \left[\frac{\alpha_I\beta_I\lambda}{1+\varphi}\right]^{\frac{1}{1-\beta_I}} \text{ and } k_x^* = \left[\frac{\alpha_x\beta_x(1+\eta)n^{\gamma}}{1+\varphi}\right]^{\frac{1}{1-\beta_x}}.$$

where φ is the shadow price of the financial constraint

$$\iota\hat{z}/2 \geq k_I\hat{z} + k_x$$

Applying Lemma 5, the firm growth rate when $\epsilon = 2$ (linear profit function) is

$$g_f(Q_f) = g_f(n\hat{z}) = \frac{F_x(k_x(n, \hat{z}))(1+\eta)}{\hat{z}} + F_I(k_I(n, \hat{z}))\lambda - \tau. \tag{4.7}$$

Disproportional Growth Rate: Deviation from Gibrat's Law

When ι is sufficiently large, $\varphi = 0$. In this case, $k_I^* = (\alpha_I\beta_I\lambda)^{\frac{1}{1-\beta_I}}$ and $F_I(k_I) = \alpha_I(\alpha_I\beta_I\lambda)^{\frac{\beta_I}{1-\beta_I}}$. Also, $k_x^* = (\alpha_x\beta_x(1+\eta)n^{\gamma})^{\frac{1}{1-\beta_x}}$ and $F_x(k_x, n) = \alpha_x(\alpha_x\beta_x(1+\eta))^{\frac{\beta_x}{1-\beta_x}} n^{-\frac{2-\beta_x}{1-\beta_x}}$

Accordingly, $g_f = \frac{C_x}{n^{\frac{2-\beta_x}{1-\beta_x}}\hat{z}} + C_I - \tau$ is decreasing in both n and $\hat{z}$, where C_x and C_I are two constants. $C_x = \alpha_x(1+\eta)(\alpha_x\beta_x(1+\eta))^{\frac{\beta_x}{1-\beta_x}}$ and $C_I = \lambda\alpha_I(\alpha_I\beta_I\lambda)^{\frac{\beta_I}{1-\beta_I}}$. Thus,

$$\frac{\partial g_f}{\partial Q} = \frac{\partial g_f}{\partial n}\frac{\partial n}{\partial Q} + \frac{\partial g_f}{\partial \hat{z}}\frac{\partial \hat{z}}{\partial Q} < 0.$$

This is the case analyzed in [AK15] where smaller firms grow faster.

Proportional Growth Rate: Gibrat's Law

Consider $\varphi > 0$.

Then $g_f = \frac{C_x}{n^{\frac{2-\beta_x}{1-\beta_x}}\hat{z}(1+\varphi)^{\frac{\beta_x}{1-\beta_x}}} + C_I/(1+\varphi)^{\frac{\beta_I}{1-\beta_I}} - \tau$. It is easy to see that $\frac{\partial g_f}{\partial \varphi} < 0$.

From the financial constraint we can see φ is defined by the implicit function

$$G(\varphi, \hat{z}, n) = \hat{z} - c_1\hat{z}(1+\varphi)^{\frac{-1}{1-\beta_I}} - c_2[n(1+\varphi)]^{\frac{-1}{1-\beta_x}} \tag{4.8}$$

where $c_1 = (\alpha_I\beta_I\lambda)^{\frac{1}{1-\beta_I}}/(0.5\iota)$ and $c_2 = (\alpha_x\beta_x(1+\eta))^{\frac{1}{1-\beta_x}}/(0.5\iota)$ are two positive constants.

It is straightforward to show that $\frac{\partial G}{\partial \varphi} = \frac{c_1}{1-\beta_I}\hat{z}(1+\varphi)^{\frac{-2+\beta_I}{1-\beta_I}} + \frac{c_2}{1-\beta_x}n^{\frac{-1}{1-\beta_x}}(1+\varphi)^{\frac{-2+\beta_x}{1-\beta_x}} > 0$ and $\frac{\partial G}{\partial n} = \frac{c_2}{1-\beta_x}(1+\varphi)^{\frac{-1}{1-\beta_x}}n^{\frac{-2+\beta_x}{1-\beta_x}}$. Also, $\frac{\partial G}{\partial \hat{z}} = 1 - c_1(1+\varphi)^{\frac{-1}{1-\beta_I}} > 0$, because $\hat{z} \geq c_1\hat{z}(1+\varphi)^{\frac{-1}{1-\beta_I}} + c_2[n(1+\varphi)]^{\frac{-1}{1-\beta_x}} > c_1\hat{z}(1+\varphi)^{\frac{-1}{1-\beta_I}}$.

By implicit function theorem,

$$\frac{\partial \varphi}{\partial \hat{z}} = -\frac{\partial G}{\partial \hat{z}}/\frac{\partial G}{\partial \varphi} < 0 \qquad \frac{\partial \varphi}{\partial n} = -\frac{\partial G}{\partial n}/\frac{\partial G}{\partial \varphi} < 0.$$

In addition,

$$\begin{aligned}
\frac{\partial g_f}{\partial \hat{z}} &= -\frac{C_x}{n^{\frac{2-\beta_x}{1-\beta_x}}\hat{z}^2(1+\varphi)^{\frac{\beta_x}{1-\beta_x}}} + \frac{\partial g_f}{\partial \varphi}\frac{\partial \varphi}{\partial \hat{z}} \\
\frac{\partial g_f}{\partial n} &= -\frac{2-\beta_x}{1-\beta_x}\frac{C_x}{n^{\frac{3-2\beta_x}{1-\beta_x}}\hat{z}(1+\varphi)^{\frac{\beta_x}{1-\beta_x}}} + \frac{\partial g_f}{\partial \varphi}\frac{\partial \varphi}{\partial n}.
\end{aligned}$$

Because the first term is negative while the second positive, the signs of $\frac{\partial g_f}{\partial \hat{z}}$ and $\frac{\partial g_f}{\partial n}$ become ambiguous. When $\frac{\partial g_f}{\partial Q} = 0$, we will return to the case [KK04] where the unconditional growth rate follows Gibrat's law: growth rate is independent of size.

4.6 Computational Details

4.6.1 Discretized Joint Distribution $h(n, \hat{z})$

(1) **Discretize** the $n - \hat{z}$ grid.

(2) Calculate the **transition matrix P**.

For each $n - \hat{z}$ pair, it can transit to

1. n and $\hat{z}\frac{1+\lambda}{1+g\Delta t}$ with probability $F_I \Delta t = \alpha_I k_I^{\beta_I} \Delta t$
2. $n+1$ and $\frac{n\hat{z}}{(n+1)(1+g\Delta t)} + \frac{q(1+\eta)}{(n+1)(1+g\Delta t)}$ with probability $nF_x\phi(q)\Delta t = n^{\gamma+1}\alpha_x k_x^{\beta_x}\phi(q)\Delta t$ where $\phi(\hat{z})$ is the marginal distribution of $\hat{z}$
3. $n-1$ and $\frac{\hat{z}}{1+g\Delta t}$ with probability $\tau\Delta t$
4. n and $\frac{\hat{z}}{1+g\Delta t}$ with probability Δt
5. n and $\hat{z}$ with probability $1 - F_I\Delta t - nF_x\Delta t - \tau\Delta t - \Delta t$.

Because updated $\hat{z}$ will not land on the grid points exactly, I use the following approximation: Suppose $a < \hat{z} < b$ where a, b are adjacent grid points. Then the probability of hitting a is $\frac{\hat{z}-a}{b-a}$ and b is $\frac{b-\hat{z}}{b-a}$. Also, the index of a, b on the grid is $\left\lfloor \frac{\log(\hat{z}/\underline{z})}{\log(1+\Delta z)} \right\rfloor$ and $\left\lceil \frac{\log(\hat{z}/\underline{z})}{\log(1+\Delta z)} \right\rceil$.

Also, boundaries are reflective.

(3) Calculate the stationary distribution of $n - \hat{z}$, **h**

$$\mathbf{h} = \mathbf{hP}.$$

h is the left eigenvector of **P**. Note that the stationary distribution of this Markov Chain exists and it is unique, since **P** is aperiodic and irreducible.

4.6.2 Value Functions

Given function inputs $n, \hat{z}, g, \tau, \phi(\hat{z})$ and Ω, we have the following iteration algorithm

$$V^{i+1}(\cdot) = \mathbf{T}_V \left\{ V^i(\cdot) \right\} \equiv V^i(\cdot) + \Delta HJB(V^i(\cdot))$$

where $\mathbf{T}_V$ describes the Bellman operator and the HJB equation is defined as

$$
\begin{aligned}
HJB[V(\cdot)] = (g-r)V(n,\hat{z}) + \max_{k_I,k_x}\{&An\hat{z}^{\epsilon-1}/\epsilon - n(\hat{z}^{\epsilon-1}k_I + k_x) \\
&+ F_I(k_I)[V(n,\hat{z}(1+\lambda)) - V(n,\hat{z})] \\
&+ n\tau[V(n-1,\hat{z}) - V(n,\hat{z})] \\
&+ nF_x(k_x,n)\left[\mathbb{E}_{\hat{z}'}V(n+1,\frac{n\hat{z}+\hat{z}'(1+\eta)}{n+1}) - V(n,\hat{z})\right] \\
&- \frac{\partial V}{\partial \hat{z}}(n,\hat{z})g\hat{z} + \varphi\left[\iota A\hat{z}^{\epsilon-1}/\epsilon - k_I\hat{z}^{\epsilon-1} - k_x\right]n\}.
\end{aligned}
$$

Neural Network Representation

Denote the state space for the neural network by $\mathbf{X}_V = \{n, \hat{z}, g, \tau, \phi(\hat{z}), \Omega\}$. Denote the respective neural network representation by $V(\mathbf{X}_V;\Theta_V)$.

Start

1. A random draw of inputs (I use 20,000 draws) $\mathbf{X}_V$
2. Initial guess of value function.

In each iteration:

1. Calculate the value function according to the last iteration's output $V^i(\mathbf{X}_V)$.
2. Calculate the results from the HJB equation (which also uses the last round's value function in the calculation) $HJB^i(\mathbf{X}_V)$.
3. Calculate the "target" for the neural network as $V^i + \Delta HJB^i$.
4. Train the neural network to fit the target.
5. The trained network is a representation of value function; call it V^{i+1}.
6. Repeat until convergence.

The neural network is calibrated with the following hyper-parameters in Table 4.7.

The machine-learning packages, TensorFlow and Keras, do most of the heavy-lifting of training the neural network. I only need to write out the discretized HJB equation.

Algorithm 1: Value Function Algorithm

Input : Uniform random draw of $\mathbf{S} = \{(\mathbf{s}_i) | \mathbf{s}_i = n, \hat{z}; \phi(\hat{z}), g, \tau, \Omega\}$

Output: Converged value functions $V(n, \hat{z}; \phi(\hat{z}), g, \tau, \Omega\}$

$i \leftarrow 0$, $\varepsilon \leftarrow 10$

while $i < max\ and\ \epsilon > tol$ **do**

  target $\leftarrow \Delta t \cdot HJB(V^i(\mathbf{S})) + V^i(\mathbf{S})$; `/* When converged,` $HJB(V^i(\mathbf{S})) = 0$ `*/`

  $V^{i+1}(\cdot) \leftarrow$ model.fit($\mathbf{S}$,target)

  $\varepsilon = \sum_{\mathbf{s}_i} |HJB(V^{i+1}(\mathbf{s}_i))|$

  $i \leftarrow i + 1$

end

Table 4.7: Neural Network Hyper Parameters

Parameters	Value
Hidden Layers	3
Hidden Units	[100,100,100]
Activation Function	Leaky ReLu
Batch Normalization	No
Gradient Descent	ADAM
Learning Rate	1e-4 to 1e-5
Batch Size	100 - 2000
Epoch	10 - 200
dt	0.07, 0.02, 0.01

CHAPTER 5

Appendix II

5.1 Selected Literature Review

This review surveys recent research on bank profitability and financial stability from theoretical, empirical, and policy perspectives. We focus on three areas that are closest to our analysis on the determinants of risks and bank profitability. The key determinants can be grouped into three main types: 1) bank business models, including the role of NII and bank characteristics; 2) cyclical conditions and structural factors, including concentration and competition; and 3) policy factors in influencing bank profitability and financial stability.

The literature on bank business models examines bank performance and risks across different business models. An important aspect of diversification in business model emphasized in the literature is the implication of NII activities, which has shown mixed evidence. Most empirical studies on U.S. banks conclude that increased reliance on NII has little to no impact on lifting bank profits and offsetting the debilitating effect of return volatilities (e.g., [DR01, Sti07, SR06, CL09, ECB15]). International studies, particularly those of European banks, paint a brighter picture and suggest a "diversification premium", implying that banks with more diversified revenue streams are more profitable, but the benefit of NII is still debatable (e.g., [BDV07, EHH10, SW11, CT10, LNR08, Koh14]).

Some recent papers propose a more nuanced and non-linear relationship between income diversification and risks. NII activities are found to be beneficial when the NII share is low ([BR16, DH10]), or when banks are small ([DDS15, AMT18]). Furthermore, based on supervisory data in Germany, [Koh14] found that diversification into NII activities were more beneficial for retail-oriented banks such as savings and cooperative banks. Using U.S. banking "call reports", [DT13] concluded that the probability of distressed bank failure declined with pure fee-based NII activities but increased with asset-based NII like investment banking and securitization. Besides bank idiosyncratic return and risks, NIIs are also shown to increase systemic risks of the banking sector ([DDS15, BDP11, EMS14]).

5.2 Proofs of Propositions and Lemmas

Proposition 3

Proof. Start from EDF. At optimal, $EDF \equiv 1 - q^* = 1 - \Phi(\frac{\mu_\pi^* + e}{\sigma_\pi^*})$. So $\frac{\partial q^*}{\partial \mu_\pi^*} = \frac{\phi(\frac{\mu_\pi^* + e}{\sigma_\pi^*})}{\sigma_\pi^*} > 0$, and thus

$$\frac{\partial EDF}{\partial \mu_\pi^*} = -\frac{\partial q^*}{\partial \mu_\pi^*} < 0$$

VaR is defined by an implicit function $F(VaR, \mu_\pi, \sigma_\pi)$, where

$F = \Phi(\frac{-VaR - \mu_\pi^* - e}{\sigma_\pi^*}) - 0.05 = 0$

To ease notation, denote $\frac{-VaR - \mu_\pi^* - e}{\sigma_\pi^*} = C_1$

$$\frac{\partial F}{\partial VaR} = -\frac{1}{\sigma_\pi^*}\phi(C_1) < 0$$
$$\frac{\partial F}{\partial \mu_\pi^*} = -\frac{1}{\sigma_\pi^*}\phi(C_1) < 0$$

By implicit function theorem,

$$\frac{\partial VaR}{\partial \mu_\pi^*} = -\frac{\frac{\partial F}{\partial \mu_\pi^*}}{\frac{\partial F}{\partial VaR}} < 0$$

■

Lemma 7. *q^* and VaR are decreasing in the standard deviation of ROA (σ_π^*) if $\mu_\pi^* + e > \epsilon$.*

$$\frac{\partial q^*}{\partial \sigma_\pi^*} < 0 \text{ and } \frac{\partial VaR}{\partial \sigma_\pi^*} < 0 \text{ if } \mu_\pi^* + e > \epsilon$$

Proof.

$$\frac{\partial q^*}{\partial \sigma_\pi^*} = -\phi(\frac{\mu_\pi^* + e}{\sigma_\pi^*})\frac{\mu_\pi^* + e}{\sigma_\pi^{*2}} < 0$$

Also

$$\frac{\partial F}{\partial \sigma_\pi^*} = \frac{VaR + \mu_\pi^* + e}{\sigma_\pi^{*2}} \phi(C_1) > 0$$

By the implicit function theorem,

$$\frac{\partial VaR}{\partial \sigma_\pi^*} = -\frac{\frac{\partial F}{\partial \sigma_\pi^*}}{\frac{\partial F}{\partial VaR}} > 0$$

■

Lemma 8. *The (expected) ROA μ_π^* is increasing in the LTA ratio l, under certain regularity conditions.*

$$\frac{\partial \mu_\pi^*}{\partial l} < 0 \textit{ if } l < \frac{1}{1+k} \textit{ and } x < 1 + \frac{(r_m - c_m)(\frac{1-\alpha}{\alpha}k - 1) + \frac{1-\alpha}{\alpha}kc_r}{r_L}$$

Proof.

$$\mu_\pi^* = \begin{cases} (1-x)r_L l + r_r(1-l)^\alpha l^{1-\alpha} - c_r(1-l) - c_f - r_D(1-e) & \text{if } l \geq \frac{1}{1+k} \\ (1-x)r_L l + (r_m - c_m)(1-l-kl) + r_r k^\alpha l - c_r kl - c_f - r_D(1-e) & \text{if } l < \frac{1}{1+k} \end{cases}$$

If $l < \frac{1}{1+k}$,

$$\frac{\partial \mu_\pi^*}{\partial l} = (1-x)r_L + (r_m - c_m)(1+k) + r_r k^\alpha - c_r k$$

where $k = \left(\frac{\alpha r_r}{c_r + r_m - c_m}\right)^{\frac{1}{1-\alpha}}$.

Note that $r_r k^\alpha - c_r k = k(r_r k^{\alpha-1} - c_r) = \frac{k}{\alpha}[c_r(1-\alpha) + r_m - c_m]$. Therefore $\frac{\partial \mu_\pi^*}{\partial l} > 0$ if and only if $(1-x)r_L + (r_m - c_m)(1+k) + \frac{k}{\alpha}[c_r(1-\alpha) + r_m - c_m] = (1-x)r_L + (r_m - c_m)(\frac{1-\alpha}{\alpha}k - 1) + \frac{1-\alpha}{\alpha}kc_r > 0$. Equivalently, when

$$x < 1 + \frac{(r_m - c_m)(\frac{1-\alpha}{\alpha}k - 1) + \frac{1-\alpha}{\alpha}kc_r}{r_L}$$

A sufficient condition will be $k > \frac{\alpha}{1-\alpha}$

■

Lemma 9. *The standard deviation ROA σ_π^* is increasing in the LTA ratio l, under certain regularity conditions.*

$$\frac{\partial \sigma_\pi^*}{\partial l} < 0 \ \ if\ l < \underline{l} = \frac{(1+k)\sigma_m^2}{(1-x)^2\sigma_L^2 + (1+k)^2\sigma_m^2 + k^{2\alpha}\sigma_r^2} < \frac{1}{1+k}$$

Proof.

$$\sigma_\pi^* = \begin{cases} (1-x)^2 l^2 \sigma_L^2 + (1-l)^{2\alpha} l^{2-2\alpha}\sigma_r^2 & \text{if } l \geq \frac{1}{1+k} \\ (1-x)^2 l^2 \sigma_L^2 + (1-l-kl)^2\sigma_m^2 + k^{2\alpha} l^2 \sigma_r^2 & \text{if } l < \frac{1}{1+k} \end{cases}$$

If $l < \frac{1}{1+k}$,

$$\frac{\partial \sigma_\pi^{2*}}{\partial l} = [2(1-x)^2\sigma_L^2 + 2(1+k)^2\sigma_m^2 + 2k^{2\alpha}\sigma_r^2]l - 2(1+k)\sigma_m^2 < 0$$

if and only if

$$l < \underline{l} = \frac{(1+k)\sigma_m^2}{(1-x)^2\sigma_L^2 + (1+k)^2\sigma_m^2 + k^{2\alpha}\sigma_r^2} = \frac{1}{\frac{(1-x)^2\sigma_L^2}{(1+k)\sigma_m^2} + (1+k) + \frac{k^{2\alpha}\sigma_r^2}{(1+k)\sigma_m^2}} < \frac{1}{1+k}$$

Therefore,

$$\frac{\partial \sigma_\pi^*}{\partial l} = \frac{1}{2\sigma_\pi^*}\frac{\partial \sigma_\pi^{2*}}{\partial l} < 0$$

∎

Proposition 4.

Proof. Based on the proof of propostion 3, Lemma 7, 8 and 9, we have

$$\frac{\partial EDF}{\partial n} = -\frac{\partial EDF}{\partial l} = \frac{\partial q^*}{\partial l} = \frac{\partial q^*}{\partial \mu_\pi^*}\frac{\partial \mu_\pi^*}{\partial l} + \frac{\partial q^*}{\partial \sigma_\pi^*}\frac{\partial \sigma_p i^*}{\partial l} > 0$$

$$\frac{\partial VaR}{\partial n} = -\frac{\partial EDF}{\partial l} = -\frac{\partial VaR}{\partial \mu_\pi^*}\frac{\partial \mu_\pi^*}{\partial l} - \frac{\partial VaR}{\partial \sigma_\pi^*}\frac{\partial \sigma_p i^*}{\partial l} > 0$$

when $x < 1 + \frac{(r_m - c_m)(\frac{1-\alpha}{\alpha}k - 1) + \frac{1-\alpha}{\alpha}kc_r}{r_L}$. Also,

$$\frac{\partial EDF}{\partial s} = \frac{\partial EDF}{\partial n}\frac{\partial n}{\partial s} > 0$$
$$\frac{\partial VaR}{\partial s} = \frac{\partial VaR}{\partial n}\frac{\partial n}{\partial s} > 0$$

∎

Corollary 6.1. *EDF and VaR are decreasing $E/A = e$ (increasing in leverage $1/e$)*

$$\frac{\partial EDF}{\partial e} < 0 \;\; \frac{\partial VaR}{\partial e} < 0$$

Proof. Similar to the proof in proposition 3

$$\frac{\partial EDF}{\partial e} = -\frac{\partial q^*}{\partial e} = -\frac{\phi(\frac{\mu_\pi^*+e}{\sigma_\pi^*})}{\sigma_\pi^*} < 0$$

and

$$\frac{\partial F}{\partial e} = -\frac{1}{\sigma_\pi^*}\phi(C_1) < 0$$

By implicit function theorem,

$$\frac{\partial VaR}{\partial e} = -\frac{\frac{\partial F}{\partial e}}{\frac{\partial F}{\partial VaR}} < 0$$

■

Proposition 5

Proof. Start from EDF. At optimal, $EDF' \equiv 1 - q^* = 1 - p * \left[\Phi(\frac{\mu_\pi^*+e}{\sigma_\pi^*}) - \Phi(\frac{\mu_\pi^*+e-\epsilon}{\sigma_\pi^*})\right] + \Phi(\frac{\mu_\pi^*+e-\epsilon}{\sigma_\pi^*})$

To ease notation, denote $B = \left[\Phi(\frac{\mu_\pi^*+e}{\sigma_\pi^*}) - \Phi(\frac{\mu_\pi^*+e-\epsilon}{\sigma_\pi^*})\right] > 0$, so

$$p^* = \frac{\epsilon + Bev}{b}$$
$$q^* = p^*B + \Phi(\frac{\mu_\pi^* + e - \epsilon}{\sigma_\pi^*})$$

Thus,

$$\frac{\partial EDF'}{\partial v} = \frac{\partial q^*}{\partial v} = -B\frac{\partial p^*}{\partial v} = -\frac{eB^2}{b} < 0$$

VaR' is defined by an implicit function $F(VaR', \mu_\pi, \sigma_\pi, v)$, where

$$F = \Phi\left(\frac{-VaR' - \mu_\pi^* - e}{\sigma_\pi^*}\right)p^* + \Phi\left(\frac{-VaR' - \mu_\pi^* - e + \epsilon}{\sigma_\pi^*}\right)(1 - p^*) - 0.05 = 0$$

Denote $\frac{-VaR'-\mu_\pi^*-e}{\sigma_\pi^*}=C_1$ and $\frac{-VaR'-\mu_\pi^*-e+\epsilon}{\sigma_\pi^*}=C_2>C_1$

$$\frac{\partial F}{\partial VaR'}=-\frac{p^*}{\sigma_\pi^*}\phi(C_1)-\frac{1-p^*}{\sigma_\pi^*}\phi(C_2)<0$$
$$\frac{\partial F}{\partial v}=-\frac{\partial p^*}{\partial v}[\Phi(C_2)-\Phi(C_1)]<0$$

By implicit function theorem,

$$\frac{\partial VaR}{\partial v}=-\frac{\frac{\partial F}{\partial v}}{\frac{\partial F}{\partial VaR'}}<0$$

■

Proposition 6

Proof. Note that

$$\frac{\partial B}{\partial e}=\frac{\left[\phi(\frac{\mu_\pi^*+e}{\sigma_\pi^*})-\phi(\frac{\mu_\pi^*+e-\epsilon}{\sigma_\pi^*})\right]}{\sigma_\pi^*}>0$$
$$\frac{\partial p^*}{\partial e}=\frac{ev}{b}\frac{\partial B}{\partial e}+\frac{Bv}{b}>0$$

and

$$\frac{\partial q^*}{\partial e}=p^*\frac{\partial B}{\partial e}+\frac{\partial p^*}{\partial e}B+\frac{\phi(\frac{\mu_\pi^*+e-\epsilon}{\sigma_\pi^*})}{\sigma_\pi^*}>0$$

Thus

$$\frac{\partial EDF'}{\partial e}=-\frac{\partial q^*}{\partial e}<0$$
$$\frac{\partial F}{\partial e}=-\frac{p^*}{\sigma_\pi^*}\phi(C_1)-\frac{1-p^*}{\sigma_\pi^*}\phi(C_2)-\frac{\partial p^*}{\partial e}[\Phi(C_2)-\Phi(C_1)]<0$$

By implicit function theorem,

$$\frac{\partial VaR'}{\partial e}=-\frac{\frac{\partial F}{\partial e}}{\frac{\partial F}{\partial VaR'}}<0$$

■

5.3 Data Sources and Definitions

5.4 Robustness Checks

Table 5.1: Bank-Specific Variables (1/2)

Data series	Definition
ROAA (%)	Return on average assets; net income as a percent of average assets
ROAE (%)	Return on average equity; net income as a percent of average equity
Price/Book (%)	Price as a percent of book value per share. Book value is calculated using common equity values and shares outstanding at the end of a financial period
Noninterest Income	Total operating income (revenues) excluding net interest income
Operating Income	Total operating income from banking, insurance, and asset management
Total Gross Loans	Loans and finance leases held for investment or held for sale; net of unearned discount and gross of loss reserves.
Total Deposits	Total deposits from customers. For U.S. banks, this is the total deposits from customers and banks.
Total Assets	All assets owned by the company as of the date indicated, as carried on the balance sheet and defined under the indicated accounting principles
Total Liabilities	Total liabilities as carried on the balance sheet and defined by the indicated accounting principles.
Total Equity	Equity as defined under the indicated accounting principles. Includes par value, paid in capital, retained earnings, and other adjustments to equity. Minority interest may be included, per relevant accounting standards.

[a] Source: S&P Global Market Intelligence's SNL database.

Table 5.2: Bank-Specific Variables (2/2)

Data series	Definition
Tier 1 Ratio (%)	Tier 1 capital ratio as defined by the latest regulatory and supervisory guidelines.
Problem Loans	The problem loan value that the company most commonly presents. If the company commonly reports multiple values, SNL selects based on the following priority (at SNL's discretion): Nonperforming Loans, Gross Impaired Loans, Net Impaired Loans, and Other Problem Loans.
Cost-to-Income (%)	Operating expense as a percent of operating income.
Cost of Funds (%)	Interest incurred on liabilities as a percent of average noninterest-bearing deposits and interest-bearing liabilities
Compensation and Benefits	Salaries, wages, bonuses, commissions, changes in reserve for future stock option expense, and other employee benefit costs.
Interest Expense	Interest on debt and other borrowings (on an incurred basis). Includes the amortization of discount (or premiums) and interest on capital leases.
Other Expense	Expense not otherwise classified.

[a.] Source: S&P Global Market Intelligence's SNL database.

Table 5.3: Financial Variables (1/2)

Data series	Source	Definition
Equity prices	Bloomberg	Daily equity price, closing price
Expected Default Frequency (EDF) (1 year)	Moody's	Forward-looking measure of actual probability of default of a bank over one year. According to the Moody's EDF model, a bank defaults when the market value of its assets falls below its liabilities payable
3-month T-bill rate	Bloomberg	Daily 3-month T-bill rate
3-month German government bond yield	Bloomberg	Daily 3-month German government bond yield
10-year T-bill rate	Bloomberg	Daily 10-year T-bill rate
10-year German government bond yield	Bloomberg	Daily 10-year German government bond yield
3-month LIBOR rate	Bloomberg	Daily 3-month LIBOR rate
3-month EURIBOR rate	Bloomberg	Daily 3-month EURIBOR rate
Credit spread of Moody's Baa-rated bonds	Bloomberg	Daily credit spread of Moody's Baa-rated bonds

Table 5.4: Financial Variables (2/2)

Data series	Source	Definition
Euro Stoxx Banks Index	Bloomberg	Daily equity index price, closing price
Euro Stoxx 50 Index	Bloomberg	Daily equity index price, closing price
MSCI Europe Real Estate Index	Bloomberg	Daily equity index price, closing price
S&P 500 Financials Index	Bloomberg	Daily equity index price, closing price
S&P 500 index returns	Bloomberg	Daily equity index price, closing price
Dow Jones U.S. Real Estate Index	Bloomberg	Daily equity index price, closing price
MSCI World Index	Bloomberg	Daily equity index price, closing price
MSCI World Financials Index	Bloomberg	Daily equity index price, closing price
MSCI World Real Estate Index	Bloomberg	Daily equity index price, closing price
VIX Index	Bloomberg	Daily equity index price, closing price

Table 5.5: Macroeconomic Variables

Data series	Source	Definition
GDP growth	IMF WEO	Growth of Gross Domestic Product, constant prices
Interest rate (3 month)	OECD	Short term (3 months) interest rate, money market
Government bond yield (10 year)	Haver Analytics	10-Year Government Bond Yield (AVG, %)
General government balance	IMF WEO	General government structural balance
Central bank claims	IMF MFS statistics, Haver Analytics	Central Bank Survey, Claims on Other Financial Corporations and Other Depository Corporations

Table 5.6: Summary Statistics

VARIABLES	(1) N	(2) mean	(3) p50	(4) sd	(5) min	(6) max
Cost of Funds (%)	4,846	1.473	1.146	1.113	0.0476	7.374
Cost-to-Income (%)	5,304	65.40	63.98	17.88	25.27	347.1
Price-to-Book Ratio (%)	4,838	139.1	128.1	68.67	12.51	515.4
ROAA (%)	5,268	0.714	0.811	0.861	-6.022	4.067
ROAE (%)	5,241	7.235	8.372	11.10	-114.7	42.03
Tier 1 Ratio (%)	5,099	13.40	12.60	4.431	4.901	50.77
Problem Loans Ratio (%)	5,134	2.797	1.375	4.603	0	40.39
Loan-to-Asset Ratio (%)	5,306	65.72	68.59	15.08	10.01	92.24
Log(Assets)	5,406	15.83	15.24	2.282	12.32	21.62
Non-Interest Income Share (%)	5,216	28.66	25.18	18.77	0.392	257.6
NII Share (%) X Loan-to-Asset Ratio (%)	5,192	1,742	1,617	1,015	28.77	14,065
Deposit-to-Liability Ratio (%)	5,377	77.17	83.84	19.82	0	99.36
Leverage Ratio	5,401	12.09	10.64	6.099	1.831	69.01
Lerner Index (%)	5,115	26.59	27.41	11.93	-58.00	56.56
ROAA/sd(ROAA) (%)	5,267	2.681	2.080	2.799	-3.063	15.47
ROAA/sd(ROAE) (%)	5,239	2.461	1.876	2.657	-3.144	15.16
Gov Structural Balance/Potential GDP (%)	5,992	-4.463	-4.258	2.645	-19.39	4.360
ST Interest Rate (%)	5,966	1.554	0.644	1.792	-0.784	15.82
Real GDP Growth Rate (%)	5,992	1.825	2.224	2.134	-9.132	25.49
Claim Growth Rate (%)	5,780	90.46	-14.16	381.5	-94.40	6,332
Delta CoVaR (95%) of Weekly Loss (%)	4,710	1.621	1.272	1.918	-4.257	10.50
VaR (95%) of Daily Loss (%)	4,722	3.118	2.569	1.901	0	12.71
Expected Default Frequency (%)	3,669	0.762	0.390	1.722	0.0198	25.15
logit(EDF)	3,669	-5.469	-5.544	0.967	-8.527	-1.091

[a.] Sources: S&P Global Market Intelligence's SNL database and IMF staff calculations.

Table 5.7: Robustness Checks for the Determinants of Risks: Static Panel

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	trEDF	trEDF	trEDF	VaR	VaR	VaR	covar	covar	covar
ROAA (%) = L	-0.214***	-0.264***	-0.284***	-0.176*	-0.467***	-0.500***	0.00250	0.0276	-0.0122
	(0.0552)	(0.0628)	(0.0635)	(0.0968)	(0.0895)	(0.0968)	(0.0414)	(0.0406)	(0.0395)
Non-Interest Income Share (%) = L	0.00893**	0.00789	0.0137***	0.0171*	-0.00701	0.0222**	0.0390***	0.0135*	0.0204***
	(0.00414)	(0.00533)	(0.00442)	(0.00915)	(0.00895)	(0.00920)	(0.00748)	(0.00766)	(0.00766)
NII Share (%) X Loan-to-Asset Ratio (%) = L	-0.000198***	-0.000174**	-0.000212***	-0.000264*	4.18e-05	-0.000275*	-0.000584***	-0.000224*	-0.000329**
	(7.23e-05)	(8.77e-05)	(8.14e-05)	(0.000160)	(0.000151)	(0.000160)	(0.000129)	(0.000130)	(0.000134)
Real GDP Growth Rate (%)	-0.0491***	-0.0562***	-0.0580***	-0.323***	-0.329***	-0.368***	-0.280***	-0.213***	-0.265***
	(0.0151)	(0.0144)	(0.0142)	(0.0332)	(0.0330)	(0.0313)	(0.0319)	(0.0286)	(0.0282)
ST Interest Rate (%)	-0.000170	0.0141	0.0169	0.0878***	0.160***	0.186***	0.0368*	-0.00173	0.0288*
	(0.0119)	(0.0122)	(0.0119)	(0.0258)	(0.0191)	(0.0184)	(0.0210)	(0.0188)	(0.0175)
Gov Structural Balance/Potential GDP (%)	-0.116***	-0.123***	-0.120***	-0.183***	-0.227***	-0.199***	-0.176***	-0.215***	-0.167***
	(0.0126)	(0.0125)	(0.0122)	(0.0191)	(0.0203)	(0.0175)	(0.0181)	(0.0176)	(0.0169)
Tier 1 Ratio (%) = L	-0.00288	0.000886	-0.00500	-0.107***	-0.0709***	-0.0851***	-0.0468***	-0.0404***	-0.0428***
	(0.0109)	(0.0109)	(0.00862)	(0.0228)	(0.0201)	(0.0192)	(0.0176)	(0.0136)	(0.0154)
Problem Loans Ratio (%) = L	0.0776***	0.0760***	0.0752***	0.110***	0.0967***	0.115***	0.00355	-0.0201**	0.00426
	(0.00857)	(0.00972)	(0.00898)	(0.0217)	(0.0197)	(0.0196)	(0.00825)	(0.00914)	(0.00804)
Lerner Index (%) = L	-0.0121***			-0.0451***			0.00905*		
	(0.00362)			(0.0116)			(0.00548)		
Deposit-to-Liability Ratio (%) = L		-0.00341			-0.0271***			-0.0291***	
		(0.00300)			(0.00668)			(0.00484)	
Leverage Ratio = L			-0.000696			0.00380			0.00858***
			(0.00251)			(0.00325)			(0.00177)
Observations	3,086	3,128	3,133	4,062	4,113	4,118	4,052	4,103	4,108

[a] The static panel regression with bank fixed effects and lagged regressors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses.

Table 5.8: Robustness Checks for the Determinants of Risks: Year FE

VARIABLES	(1) covar	(2) covar	(3) VaR	(4) VaR	(5) trEDF	(6) trEDF
ROAA (%)	-0.188**		-0.580***		-0.142**	
	(0.0926)		(0.0913)		(0.0631)	
Price-to-Book Ratio (%)		0.00120		-0.00252***		-0.000487
		(0.00118)		(0.000841)		(0.000552)
Non-Interest Income Share (%)	0.0273***	0.0214***	0.0108*	0.00507	0.00539**	0.00684**
	(0.00812)	(0.00646)	(0.00650)	(0.00545)	(0.00267)	(0.00270)
NII Share (%) X Loan-to-Asset Ratio (%)	-0.000347**	-0.000335**	-0.000263**	-0.000176*	-9.56e-05**	-0.000105**
	(0.000159)	(0.000142)	(0.000113)	(9.67e-05)	(4.78e-05)	(4.11e-05)
Tier 1 Ratio (%)	-0.0411**	-0.0557***	-0.0377***	-0.0587***	0.0108	0.00429
	(0.0189)	(0.0190)	(0.0135)	(0.0140)	(0.00733)	(0.00674)
Problem Loans Ratio (%)	-0.00559	-0.00516	0.0551***	0.0581***	0.0292***	0.0325***
	(0.00811)	(0.00795)	(0.0139)	(0.0124)	(0.00461)	(0.00541)
Real GDP Growth Rate (%)	-0.0293	-0.0379	-0.0772***	-0.113***	-0.0221**	-0.0323***
	(0.0296)	(0.0320)	(0.0288)	(0.0283)	(0.00973)	(0.0111)
ST Interest Rate (%)	0.0472	0.0123	0.0115	-0.0372	0.165***	0.157***
	(0.0764)	(0.0655)	(0.0513)	(0.0485)	(0.0224)	(0.0214)
Gov Structural Balance/Potential GDP (%)	-0.0602***	-0.0265	-0.0775***	-0.0614***	-0.0317***	-0.0280***
	(0.0225)	(0.0182)	(0.0181)	(0.0173)	(0.00847)	(0.00864)
Lerner Index (%)	0.0414***	0.0259***	-0.00271	-0.0168**	0.00345	0.00645**
	(0.00954)	(0.00681)	(0.00864)	(0.00790)	(0.00345)	(0.00302)
Delta CoVaR (95%) of Weekly Loss (%) = L	0.182***	0.368***				
	(0.0218)	(0.0343)				
VaR (95%) of Daily Loss (%) = L			0.458***	0.498***		
			(0.0399)	(0.0389)		
logit(EDF) = L					0.664***	0.699***
					(0.0301)	(0.0332)
Observations	3,864	3,830	3,867	3,833	2,971	2,943
Lags	2 - 6	3 - 9	2 - 7	2 - 7	3 - 5	3 - 6
No. of Intruments	380	412	421	420	260	309
No. of Banks	386	385	387	386	290	287
Hansen p-Value	0.371	0.667	0.762	0.786	0.127	0.681
AR2 p-Value	0.0166	6.46e-05	0.0796	0.560	0.0311	6.77e-10
AR3 p-Value	0.788	0.761	0.543	0.321	0.138	0.000174
Year Fixed Effects	Y	Y	Y	Y	Y	Y

[a.] The dynamic panel regression with year fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses.

REFERENCES

[AA90] Zoltan J. Acs and David B. Audretsch. *Innovation and Small Firms.* The MIT Press, Cambridge, Massachusetts, 1990.

[AAA17] Daron Acemoglu, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr. "Innovation, Reallocation, and Growth." *Ssrn*, 2017.

[AAB10] Philippe Aghion, George Marios Angeletos, Abhijit Banerjee, and Kalina Manova. "Volatility and growth: Credit constraints and the composition of investment." *Journal of Monetary Economics*, **57**(3):246–265, 2010.

[AAB12] Philippe Aghion, Philippe Askenazy, Nicolas Berman, Gilbert Cette, and Laurent Eymard. "Credit constraints and the cyclicality of R&D investment: Evidence from France." *Journal of the European Economic Association*, **10**(5):1001–1024, 2012.

[AAH14] Philippe Aghion, Ufuk Akcigit, and Peter Howitt. "What Do We Learn From Schumpeterian Growth Theory?" *Handbook of Economic Growth*, **2**:515–563, 2014.

[AAM17] A Continuous-time Approach, Yves Achdou, and Benjamin Moll. "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach." 2017.

[AB95] Manuel Arellano and Olympia Bover. "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics*, **68**(1):29–51, 1995.

[AB16] Tobias Adrian and Markus K. Brunnermeier. "CoVaR." *American Economic Review*, **106**(7):1705–1741, 2016.

[AC15] Daron Acemoglu and Dan Cao. "Innovation by entrants and incumbents." *Journal of Economic Theory*, **157**:255–294, 2015.

[ACG16] Diego Anzoategui, Diego Comin, Mark Gertler, and Joseba Martinez. "Endogenous Technology Adoption and R&D as Sources of Business Cycle Persistence." *NBER Working Paper*, pp. 1–54, 2016.

[ADZ14] Deniz Anginer, Asli Demirguc-Kunt, and Min Zhu. "How does competition affect bank systemic risk?" *Journal of Financial Intermediation*, **23**(1):1–26, 2014.

[AE06] David B. Audretsch and Julie Ann Elston. "Can institutional change impact high-technology firm growth?: Evidence from Germany's Neuer Markt." *Journal of Productivity Analysis*, **25**(1):9–23, 2006.

[AE10] David B. Audretsch and Julie Ann Elston. "On the relationship between firm size and growth: Have we been asking the wrong questions?" 2010.

[AHS13] Ufuk Akcigit, Douglas Hanley, and Nicolas Andre Benigno Serrano-Velarde. "Back to Basics: Basic Research Spillovers, Innovation Policy and Growth." *Ssrn*, (19473):1–57, 2013.

[AK15] Ufuk Akcigit and William R. Kerr. "Growth Through Heterogeneous Innovations." *Ssrn*, **16443**:1–80, 2015.

[Akc08] Ufuk Akcigit. "Firm Size, Innovation Dynamics and Growth." *Manuscript*, pp. 1–67, 2008.

[AMT18] Pejman Abedifar, Philip Molyneux, and Amine Tarazi. "Non-interest income and bank lending." *Journal of Banking and Finance*, **87**:411–426, 2018.

[ASZ13] Mario Daniele Amore, Cdric Schneider, and Alminas Žaldokas. "Credit supply and corporate innovation." *Journal of Financial Economics*, **109**(3):835–855, 2013.

[Bar07] Gadi Barlevy. "On the cyclicality of research and development." *American Economic Review*, **97**(4):1131–1164, 2007.

[BB98] Richard Blundell and Stephen Bond. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, **87**(1):115–143, 1998.

[BCH17] Sylvain Benoit, Jean Edouard Colliard, Christophe Hurlin, and Christophe Pérignon. "Where the risks lie: A survey on systemic risk." *Review of Finance*, **21**(1):109–152, 2017.

[BD05] John H. Boyd and Gianni De Nicoló. "The theory of bank risk taking and competition revisited." *Journal of Finance*, **60**(3):1329–1343, 2005.

[BD06] Thorsten Beck and Asli Demirguc-Kunt. "Small and medium-size enterprises: Access to finance as a growth constraint." *Journal of Banking and Finance*, **30**(11):2931–2943, 2006.

[BDL08] Thorsten Beck, Asli Demirguc-Kunt, Luc Laeven, and Ross Levine. "Finance, firm size, and growth." *Journal of Money, Credit and Banking*, **40**(7):1379–1405, 2008.

[BDM05] Thorsten Beck, Asli Demirgüç-Kunt, and Vojislav Maksimovic. "Financial and legal constraints to growth: Does firm size matter?", 2005.

[BDP11] Markus K. Brunnermeier, G. Nathan Dong, and Darius Palia. "Banks Non-Interest Income and Systemic Risk." *Ssrn*, 2011.

[BDS13] Thorsten Beck, Olivier De Jonghe, and Glenn Schepens. "Bank competition and stability: Cross-country heterogeneity." *Journal of Financial Intermediation*, **22**(2):218–244, 2013.

[BDV07] Lieven Baele, Olivier De Jonghe, and Rudi Vander Vennet. "Does the stock market value bank diversification?" *Journal of Banking and Finance*, **31**(7):1999–2023, 2007.

[BFP09] James R. Brown, Steven M. Fazzari, and Bruce C. Petersen. "Financing innovation and growth: Cash flow, external equity, and the 1990s r&d boom." *Journal of Finance*, **64**(1):151–185, 2009.

[BGH17] Claudio Borio, Leonardo Gambacorta, and Boris Hofmann. "The influence of monetary policy on bank profitability." *International Finance*, **20**(1):48–63, 2017.

[BKM19] Francesco Bianchi, Howard Kung, and Gonzalo Morales. "Growth, slowdowns, and recoveries." *Journal of Monetary Economics*, **101**:47–63, 2019.

[BKT09] Allen N. Berger, Leora F. Klapper, and Rima Turk-Ariss. "Bank competition and financial stability." *Journal of Financial Services Research*, **35**(2):99–118, 2009.

[BM06] S. Bond and C. Meghir. "Dynamic Investment Models and the Firm's Financial Policy." *The Review of Economic Studies*, **61**(2):197–222, 2006.

[BN05] John HO Boyd and Gianni D E Nicol. "The Theory of Bank Risk Taking." *Journal of Finance*, **60**(3):1329–1343, 2005.

[BP11] James R. Brown and Bruce C. Petersen. "Cash holdings and R&D smoothing." *Journal of Corporate Finance*, **17**(3):694–709, 2011.

[BR16] Arnoud W.A. Boot and Lev Ratnovski. "Banking and trading." *Review of Finance*, **20**(6):2219–2246, 2016.

[BSF06] J A Bikker, Laura Spierdijk, and Paul Finnie. "The impact of bank size on market power." 2006.

[BT02] Leonardo Becchetti and Giovanni Trovato. "The determinants of growth for small and medium sized firms. The role of the availability of external finance." *Small Business Economics*, **19**(4):291–306, 2002.

[BT10] David Besanko and Anjan V. Thakor. "Relationship banking, deposit insurance and bank portfolio choice." In *Capital markets and financial intermediation*, pp. 292–319. 2010.

[Cav98] Richard E Caves. "Industrial organization and new findings on the turnover and mobility of firms." *Journal of Economic Literature*, 1998.

[CCW17] Xin (Simba) Chang, Yangyang Chen, Sarah Qian Wang, Kuo Zhang, and Wenrui Zhang. "Credit Default Swaps and Corporate Innovation." 2017.

[CG06] Diego Comin and Mark Gertler. "Medium-term business cycles." *American Economic Review*, **96**(3):523–551, 2006.

[Cho10] Byeongyong Paul Choi. "The U.S. Property and Liability insurance industry: Firm growth, size, and age." *Risk Management and Insurance Review*, **13**(2):207–224, 2010.

[CL09] Christian Calmès and Ying Liu. "Financial structure change and banking income: A Canada-U.S. comparison." *Journal of International Financial Markets, Institutions and Money*, **19**(1):128–139, 2009.

[CMT15] Jess Cornaggia, Yifei Mao, Xuan Tian, and Brian Wolfe. "Does banking competition affect innovation?" *Journal of Financial Economics*, **115**(1):189–209, 2015.

[Coa09] Alex Coad. *The Growth of Firms - a Survey of Theories and Empirical Evidence*. Edward Elgar Publishing, 2009.

[Coh10] Wesley M. Cohen. "Fifty years of empirical studies of innovative activity and performance." In *Handbook of the Economics of Innovation*, volume 1, pp. 129–213. 2010.

[COS13] Sudheer Chava, Alexander Oettl, Ajay Subramanian, and Krishnamurthy V. Subramanian. "Banking deregulation and innovation." *Journal of Financial Economics*, **109**(3):759–774, 2013.

[CP02] Robert E. Carpenter and Bruce C. Petersen. "Is the growth of small firms constrained by internal finance?", 2002.

[CT10] Christian Calmès and Raymond Théoret. "The impact of off-balance-sheet activities on banks returns: An application of the ARCH-M to Canadian data." *Journal of Banking and Finance*, **34**(7):1719–1728, 2010.

[DDS15] Olivier De Jonghe, Maaike Diepstraten, and Glenn Schepens. "Banks size, scope and systemic risk: What role for conflicts of interest?" *Journal of Banking and Finance*, **61**:S3–S13, 2015.

[De 16] Maarten De Ridder. "Investment in Productivity and the Long-Run Effect of Financial Crises on Output." 2016.

[DH10] Asli Demirgüç-Kunt and Harry Huizinga. "Bank activity and funding strategies: The impact on risk and returns." *Journal of Financial Economics*, **98**(3):626–650, 2010.

[DR01] Robert Deyoung and Karin P. Roland. "Product Mix and Earnings Volatility at Commercial Banks: Evidence from a Degree of Total Leverage Model." *Journal of Financial Intermediation*, **10**(1):54–84, 2001.

[DT13] Robert DeYoung and Gkhan Torna. "Nontraditional banking activities and bank failures during the financial crisis." *Journal of Financial Intermediation*, **22**(3):397–421, 2013.

[Dua18] Victor Duarte. "Machine Learning for Continuous-Time Economics." 2018.

[DY14] Francis X. Diebold and Kamil Yilmaz. "On the network topology of variance decompositions: Measuring the connectedness of financial firms." In *Journal of Econometrics*, volume 182, pp. 119–134, 2014.

[ECB15] ECB. "Bank profitability challenges in euro area banks: the role of cyclical and structural factors." *ECB Financial Stability Review*, **2015**(May):134–145, 2015.

[EHH10] Ralf Elsas, Andreas Hackethal, and Markus Holzhäuser. "The anatomy of bank diversification." *Journal of Banking and Finance*, **34**(6):1274–1287, 2010.

[EMS14] Robert F. Engle, Fariborz Moshirian, Sidharth Sahgal, and Bohui Zhang. "Banks Non-Interest Income and Global Financial Stability." *Ssrn*, **015**(April 2014), 2014.

[FDA04] Yoshi Fujiwara, Corrado Di Guilmi, Hideaki Aoyama, Mauro Gallegati, and Wataru Souma. "Do Pareto-Zipf and Gibrat laws hold true? An analysis with European firms." *Physica A: Statistical Mechanics and its Applications*, **335**(1-2):197–216, 2004.

[FHN18] Jess Fernández-Villaverde, Samuel Hurtado, and Galo Nuño. "Financial Frictions and the Wealth Distribution." 2018.

[FHP06] Steven M. Fazzari, R. Glenn Hubbard, Bruce C. Petersen, Alan S. Blinder, and James M. Poterba. "Financing Constraints and Corporate Investment." *Brookings Papers on Economic Activity*, **1988**(1):141, 2006.

[FM00] Jos C. Fariñas and Lourdes Moreno. "Firms' growth, size and age: A nonparametric approach." *Review of Industrial Organization*, **17**(3):249–265, 2000.

[FR08] Xavier Freixas and Jean-Charles Rochet. *Summary for Policymakers*. 2008.

[Gar16] Daniel Garcia-Macia. "The Financing of Ideas and the Great Deviation." *Ssrn*, 2016.

[HL10] Bronwyn H. Hall and Josh Lerner. "The financing of R&D and innovation." *Handbook of the Economics of Innovation*, **1**(1 C):609–639, 2010.

[Int17] International Monetary Fund. "Spain: Financial Sector Assessment Program-Technical Note-Determinants of Bank Profitability." Technical Report 339, International Monetary Fund, Washington, DC., 2017.

[JLS13] Gabriel Jiménez, Jose A. Lopez, and Jess Saurina. "How does competition affect bank risk-taking?" *Journal of Financial Stability*, 2013.

[JQ12] Urban Jermann and Vincenzo Quadrini. "Macroeconomic effects of financial shocks.", 2012.

[Kab19] Zeynep Kabukcuoglu. "The cyclical behavior of R&D investment during the Great Recession." *Empirical Economics*, **56**(1):301–323, 2019.

[Kee90] Michael C. Keeley. "Deposit insurance, risk, and market power in banking." *American Economic Review*, 1990.

[KK04] TorJakob Klette and Samuel Kortum. "Innovating Firms and Aggregate Innovation." *Journal of Political Economy*, **112**(5):986–1018, 2004.

[KLC16] Dong Hyeon Kim, Shu Chin Lin, and Ting Cih Chen. "Financial structure, firm size and industry growth." *International Review of Economics and Finance*, **41**:23–39, 2016.

[KN15] William R Kerr and Ramana Nanda. "Financing Innovation.", 2015.

[Koh14] Matthias Köhler. "Does non-interest income make banks more risky? Retail- versus investment-oriented banks." *Review of Financial Economics*, **23**(4):182–193, 2014.

[LM08] Rasmus Lentz and Dale T. Mortensen. "An Empirical Model of Growth Through Product Innovation." *Econometrica*, **76**(6):1317–1373, 2008.

[LMR12] Germn López-Espinosa, Antonio Moreno, Antonio Rubia, and Laura Valderrama. "Short-term wholesale funding and systemic risk: A global CoVaR approach." *Journal of Banking and Finance*, 2012.

[LNR08] Laetitia Lepetit, Emmanuelle Nys, Philippe Rous, and Amine Tarazi. "Bank income structure and risk: An empirical analysis of European banks." *Journal of Banking and Finance*, **32**(8):1452–1467, 2008.

[LSN10] J. Leitão, Z. Serrasqueiro, and P.M. Nunes. "Testing Gibrat's law for listed Portuguese companies: A quantile approach." *International Research Journal of Finance and Economics*, **37**:147–158, 2010.

[LV13] Luc Laeven and Fabin Valencia. "The Real Effects of Financial Sector Interventions during Crises." *Journal of Money, Credit and Banking*, **45**(1):147–177, 2013.

[LVS05] Robert Lensink, Paul Van Steen, and Elmer Sterken. "Uncertainty and growth of the firm." *Small Business Economics*, **24**(4):381–391, 2005.

[Man18] William Mann. "Creditor rights and innovation: Evidence from patent collateral." *Journal of Financial Economics*, **130**(1):25–47, 2018.

[MM84] Stewart C. Myers and Nicholas S. Majluf. "Corporate financing and investment decisions when firms have information that investors do not have." *Journal of Financial Economics*, **13**(2):187–221, 1984.

[MNP18] Ben S. Meiselman, Stefan Nagel, and Amiyatosh K. Purnanandam. "Judging Banks' Risk by the Profits They Report." 2018.

[Mol14] Benjamin Moll. "Productivity losses from financial frictions: Can self-financing undo capital misallocation?" *American Economic Review*, **104**(10):3186–3221, 2014.

[MRV16] Natalya Martynova, Lev Ratnovski, and Razvan Vlahu. "Bank Profitability and Risk-Taking." *IMF Working Papers*, 2016.

[MV96] Carmen Matutes and Xavier Vives. "Competition for deposits, fragility, and insurance." *Journal of Financial Intermediation*, 1996.

[MV00] Carmen Matutes and Xavier Vives. "Imperfect competition, risk taking, and regulation in banking." *European Economic Review*, **44**(1):1–34, 2000.

[MX14] Virgiliu Midrigan and Daniel Yi Xu. "Finance and misallocation: Evidence from plant-level data." *American Economic Review*, **104**(2):422–458, 2014.

[MX17] Sheheryar Malik and TengTeng Xu. "Interconnectedness of Global Systemically-Important Banks and Insurers." 2017.

[Mye06] Stewart C. Myers. "The Capital Structure Puzzle." *The Journal of Finance*, **39**(3):575, 2006.

[NAA14] Islam A. Nassar, Mahmoud Khalid Almsafir, and Maher H. Al-Mahrouq. "The Validity of Gibrat's Law in Developed and Developing Countries (20082013): Comparison based Assessment." *Procedia - Social and Behavioral Sciences*, **129**:266–273, 2014.

[NN14] Ramana Nanda and Tom Nicholas. "Did bank distress stifle innovation during the Great Depression?" *Journal of Financial Economics*, **114**(2):273–292, 2014.

[Pei16] Oana Peia. "Banking crises and investments in innovation." 2016.

[Rep04] Rafael Repullo. "Capital requirements, market power, and risk-taking in banking." *Journal of Financial Intermediation*, **13**(2):156–182, 2004.

[Roo09] David Roodman. "How to do xtabond2: An introduction to difference and system GMM in Stata." *Stata Journal*, **9**(1):86–136, 2009.

[RZ98] Raghuram G. Rajan and Luigi Zingales. "Financial Dependence and Growth." *American Economic Review*, **88**(3):559–586, 1998.

[Sch17] Tom Schmitz. "Endogenous Growth, Firm Heterogeneity and the Long-run Impact of Financial Crises." 2017.

[SKT06] Enrico Santarelli, Luuk Klomp, and A. Roy Thurik. "Gibrats Law: An Overview of the Empirical Literature." In *Entrepreneurship, Growth, and Innovation*. 2006.

[SR06] Kevin J. Stiroh and Adrienne Rumble. "The dark side of diversification: The case of US financial holding companies." *Journal of Banking and Finance*, **30**(8):2131–2161, 2006.

[Sti02] Kevin J. Stiroh. "Diversification in Banking: Is Noninterest Income the Answer?" *Ssrn*, 2002.

[Sti07] Kevin J. Stiroh. "A Portfolio View of Banking with Interest and Noninterest Activities." *Journal of Money, Credit, and Banking*, **38**(5):1351–1361, 2007.

[Sut97] John Sutton. "Gibrat's Legacy." *Journal of Economic Literature*, 1997.

[SV07] Enrico Santarelli and Marco Vivarelli. "Entrepreneurship and the process of firms' entry, survival and growth." *Industrial and Corporate Change*, **16**(3):455–488, 2007.

[SV10] Andrei Shleifer and Robert W. Vishny. "Unstable banking." *Journal of Financial Economics*, **97**(3):306–318, 2010.

[SW81] JE Stiglitz and Andrew Weiss. "Credit rationing in markets with imperfect information." *The American economic review*, 1981.

[SW11] Sarah Sanya and Simon Wolfe. "Can Banks in Emerging Economies Benefit from Revenue Diversification?" *Journal of Financial Services Research*, **40**(1):79–101, 2011.

[Van05] Johannes Van Biesebroeck. "Firm Size Matters: Growth and Productivity Growth in African Manufacturing." *Economic Development and Cultural Change*, **53**(3):545–583, 2005.