



**HAL**  
open science

# Essays on Corporate Ownership and Human Capital

Camille Hebert

► **To cite this version:**

Camille Hebert. Essays on Corporate Ownership and Human Capital. Business administration. Université Paris sciences et lettres; Tilburg School of Economics and Management (TISEM), 2019. English. NNT : 2019PSLED073 . tel-03220171

**HAL Id: tel-03220171**

**<https://theses.hal.science/tel-03220171v1>**

Submitted on 7 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à l'Université Paris-Dauphine  
Dans le cadre d'une cotutelle avec Tilburg University

**Essais sur la Structure du Capital et le Capital Humain**  
Essays on Corporate Ownership and Human Capital

Soutenue par

**Camille HEBERT**

Le 11 octobre 2019

Ecole doctorale n° ED 543

**Ecole doctorale de Dauphine**

Spécialité

**Gestion, Finance**

Composition du jury :

Pascal DUMONTIER Professeur, Université Paris Dauphine	<i>Président</i>
Jasmin GIDER Professeur Assistant, Tilburg University	<i>Rapporteur</i>
Christoph SCHNEIDER Professeur Assistant, Tilburg University	<i>Rapporteur</i>
David ROBINSON Professeur, Duke University	<i>Examineur</i>
Oliver SPALT Professeur, University of Mannheim	<i>Examineur</i>
Edith GINGLINGER Professeur, Université Paris Dauphine	<i>Directrice de thèse</i>
Luc RENNEBOOG Professeur, Tilburg University	<i>Directeur de thèse</i>

# ESSAYS ON CORPORATE OWNERSHIP AND HUMAN CAPITAL

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan Tilburg University op gezag van prof. dr. G.M. Duijsters, als tijdelijk waarnemer van de functie rector magnificus en uit dien hoofde vervangend voorzitter van het College voor promoties, en Université Paris-Dauphine op gezag van de president, prof. I. Huault, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de portrettenzaal van Tilburg University  
op vrijdag 11 oktober 2019 om 10.00 uur door

CAMILLE GINETTE ANNICK HÉBERT

geboren op 31 augustus 1991 te Parijs, Frankrijk.

PROMOTORES: Prof. dr. E. Ginglinger  
Prof. dr. L.D.R. Renneboog

OVERIGE COMMISSIELEDEN: Prof. dr. P. Dumontier  
Dr. J. Gider  
Prof. dr. D.T. Robinson  
Dr. C.A.R. Schneider  
Prof. dr. O.G. Spalt

# Acknowledgements

*“The study of economics does not seem to require any specialized gifts of an unusually high order. Is it not, intellectually regarded, a very easy subject compared with the higher branches of philosophy and pure science? Yet good, or even competent, economists are the rarest of birds. An easy subject, at which very few excel! The paradox finds its explanation, perhaps, in that the master-economist must possess a rare combination of gifts. He must reach a high standard in several different directions and must combine talents not often found together. He must be mathematician, historian, statesman, philosopher in some degree. He must understand symbols and speak in words. He must contemplate the particular in terms of the general, and touch abstract and concrete in the same flight of thought. He must study the present in the light of the past for the purposes of the future. No part of man’s nature or his institutions must lie entirely outside his regard. He must be purposeful and disinterested in a simultaneous mood; as aloof and incorruptible as an artist, yet sometimes as near the earth as a politician.”*

— John Maynard Keynes in his essay on Alfred Marshall.

Writing this dissertation has been a tremendous learning experience and a rewarding chapter in my professional and personal development. It would not have been possible without the support and guidance of many different people. Now is the time to thank them!

First, my gratitude goes to my advisors, Edith Ginglinger and Luc Renneboog, who have been outstanding mentors and co-authors to me. I am thankful to Edith for her attentive guidance, her uprightness and the freedom to pursue my research interests. I am grateful to Luc for his academic dedication, his intellectual perseverance, for his contagious enthusiasm and showing me that the knowledge of a well-rounded academic goes beyond financial economics. I am forever indebted to Luc and Edith for their support and faith in me.

Next, I warmly thank Oliver Spalt, Manju Puri and David Robinson whose

---

research papers shaped the way I think about my work and helped me raise the ambition of my papers every time we met. Oliver pushed me to think about the big picture and helped me raise the bar. Manju offered feedback on my work from a very early stage and invited me to spend six intellectually rich months at Duke University. My visit to Duke was a turning point in my development as a young researcher and has opened new horizons for me. David was always available to discuss ideas, introduced me to the entrepreneurial finance research world, and kindly reoriented my work when it was needed.

I would also like to express my gratitude to Pascal Dumontier, Jasmin Gider, and Christoph Schneider for accepting to take part in my dissertation committee. I have appreciated a lot the time and effort they have spent reading my chapters and their comments are very helpful for my current and future work.

I also thank the professors at my two institutions who educated me about finance and with whom I had the chance to interact with. In particular, I would like to thank Marie-Pierre Dargnies, Marie-Agnès Leutenegger, Tamara Nefedova, Marius Zoican at Dauphine, and Lieven Baele, Fabio Braggion, Stefano Cassela, Frank DeJongue, Joost Driessen at Tilburg. Thank you also to Françoise Carbon at Dauphine, Loes de Groot, Helma Tigchelaar and Marie-Cécile Kwinten at Tilburg, who behind the scene make sure everything works. A special thank to Loes for her tremendous help during the job market period.

I am grateful to the French State for providing free education and for funding my doctoral studies. I also thank the Université Paris-Dauphine and Tilburg University for financing several conferences and my visiting at Duke University. I thank the American Finance Association, the Financial Intermediation Research Society, the Financial Research Association, the Mitsui Financial Center, the NBER and Kauffman Foundation, the University of Colorado Boulder and the Western Finance Association for including Ph.D. papers in their conference programs and offering financial support to participate in the conference. On another note, I would like to thank my students in the Master Contrôle Compatibilité Audit for offering a breath from research every week in the past five years and reminding me that we do research to ultimately transmit knowledge.

I have been lucky enough to meet outside Dauphine and Tilburg, professors who

took time to provide feedback on my work, give me tips and above all for showing me that academia can be a generous place to be. To name a few, I would like to thank Manuel Adelino, Francois Derrien, Zsuzsanna Fluck, Laurent Frésard, John Graham, Denis Gromb, Johan Hombert, Roni Michaely, Paige Ouimet, Gordon Phillips and Geoff Tate. I also thank young assistant professors who have been role models for me and helped me in various ways: Sylvain Catherine, Paul Karehnke, Elisabeth Kempf and Adrien Matray.

I warmly thank my co-authors and friends Paul Beaumont and Victor Lyonnet. When we started to work together during my second year of Ph.D., I had no idea how constructive and rewarding will be our collaboration. Since then, we have grown together as young researchers and have written together one paper and half, and counting! I cannot stress enough how much I have learned by working with Paul and Victor and how much our “not so productive” discussions have kept me motivated. I do hope to pursue this fruitful collaboration in the coming years.

During the five years of my Ph.D., I had the chance to meet other fellow Ph.D. students who made the Ph.D. a fun and enjoyable experience on a daily basis. I thank Arthur Beddock and David Zerbib for accepting to be my paranymphs. I thank Marion Boisseau-Sierra, Mario Milone, Yasmine Skalli and Eric Thorez for our chats and more philosophical discussions; Peter Brok, Maaïke Diepstraten, Marshall Ma, Ekaterina Neretina, Gabor Neszveda, Emanuele Rizzo, Cara Vansteensteen and Haikun Zhu for their friendly companionship at Tilburg; Ahmet Delergi, Valentina Giannini, Tatiana Lluent and Margaux Luffade for their warm welcome at Duke, as well as the other PhD students I met along the way.

Moving forward, I would like to thank my new colleagues at the University of Toronto, Pat Akey, Claire Célérier, Alexandre Corhay, Peter Cziraki, Nicolas Inostroza, Charles Martineau, Irene Yi to name a few, for having me here and for our fruitful interactions to come.

Finally, I thank my mother, Sylvette Pasquet, and my little brother, Thibaud Hebert, for their unconditional love, encouragement and support along all these years.

Thank you, everyone!

Toronto, August 2019.

---

*“La lutte elle-même vers les sommets suffit à remplir un cœur d’homme.  
Il faut imaginer Sisyphe heureux.”*

— Albert Camus in *The Mythe of Sisyphe*.

# Introduction

This Ph.D. dissertation consists of three chapters that constitute independent research articles. Each chapter of this thesis focuses on the firm's organizational structure at a different stage of its life cycle: early-stage, growth, business group. The first chapter investigates the underlying reasons of the gender funding gap in the venture capital industry. It highlights a significant role for investors' stereotypes that ultimately impedes minority-founded startups' growth. The second chapter identifies conditions under which firms choose to grow by buying an already existing company as opposed to building on their in-house resources. The third chapter focuses on large business groups and adds evidence that investors are not always aware of the boundaries of the firm and miss predictive information released at another level of the group.

The first chapter shows that external equity stereotypes impede start-ups' development. I use detailed administrative data on the population of start-ups in France. I find that female-founded start-ups are on average 20% less likely to contract with venture capitalists, angel investors, and with other private equity investors. However, I find that this is no longer the case in female-dominated industries. Female-founded start-ups in female-dominated sectors are equally to more likely to raise equity relative to male entrepreneurs in female-dominated sectors. Female-entrepreneurs in female-dominated sectors are also more likely to use external equity financing than women entrepreneurs in male-dominated sectors. The novelty of this paper is to show that women are not always at a disadvantage when it comes to contracting with financiers: female entrepreneurs are perceived better at female activities, and male entrepreneurs are perceived better at male activities. Because of context-dependent stereotypes, investors over-estimate the entrepreneurial abilities of individuals who fit the representative gender of an activity (Bordalo et al.,

---

2016). Furthermore, I show that conditional on raising external equity financing, minority-owned start-ups perform better and grow faster than start-ups founded by an entrepreneur who belongs to the representative group within and across sectors. Indeed, successfully equity-funded female-founded start-ups in male-dominated sectors and male-founded start-ups in female-dominated sectors outperform on average. The evidence suggests that the bar to raise external equity for minority-owned start-ups was set higher than for non-minority owned start-ups. This study has implications for designing optimal policies targeting young firms. It suggests that programs helping minorities to enter into environments in which they are underrepresented could be useful to attenuate the effects of context-dependent stereotypes.

The second chapter proposes a resource-based explanation to explain firms' diversification choices. A firm can grow and enter into a new product market by building on existing in-house resources (internal entry) or by acquiring a company already operating in this market (external entry). We propose a model of production with endogenous choice of the workforce to draw empirical predictions of diversification choices and to derive an original measure of firm-level human capital. The measure of the firm's internal human capital captures to which extent a firm's existing workforce is adapted to the sector of entry. We build a novel dataset which links M&A deals from standard commercial databases to the French matched employer-employee dataset and a dataset of the detailed breakdown of firms' sales. The entry is a "buy" if the entity that starts selling in the new sector was recently acquired by the firm, whereas a firm "builds" if the entry is made through an existing subsidiary. We document that 2% (10%, weighted by entry sales) of firms enter a new sector through the acquisition of a target firm. We show that within the same sector of origin  $\times$  entry  $\times$  year, firms tend to diversify by acquisition when their human capital is not adapted to the sector of entry, and especially when the local labor market for key occupations to operate in the sector of entry is tight. Taken together, these findings suggest that labor market frictions are crucial to understanding the role of human capital in the decision to build or buy. In addition, this paper sheds light on internal entries – also called organic growth – as a useful counterfactual to think about acquisitions and takeovers.

Although the two first chapters focus on firms' growth and ultimately the form-

ing of business groups, they do not explicitly address the question of value creation implied by these types of firms. The third chapter of this thesis asks whether investors are aware of the boundaries of the firm and price information released at another level of the business group. Using an international sample of business groups which comprehend at least two publicly listed entities connected through significant ownership and control links, our main tests consist of regressing parent companies' abnormal returns on subsidiaries' earnings announcements and subsidiaries' abnormal returns on parent companies' earnings announcements, respectively. We find that parent companies' investors do not always react on time – or with delay – to their subsidiaries' earnings announcements. We also find that subsidiaries' investors on average do not react to parents' earnings announcements, although these events convey predictive information about what will be announced later at the subsidiary's level. This finding suggests that subsidiaries' investors are mostly unaware of the ownership connection with the parent company. Finally, we exploit cross-sectional variations in the salience of parent-subsidiary ownership connections and investors' degree of investors' sophistication.

# Contents

<b>1</b>	<b>Mind the Gap: Gender Stereotypes and Entrepreneur Financing</b>	<b>14</b>
1.1	Introduction . . . . .	16
1.2	Theoretical Framework . . . . .	24
1.2.1	Set-up . . . . .	24
1.2.2	Discrimination and funding error . . . . .	25
1.2.3	Taste-based discrimination . . . . .	26
1.2.4	Statistical discrimination . . . . .	27
1.2.5	Discrimination with stereotypes . . . . .	29
1.2.6	Aggregate effects and policy implications . . . . .	32
1.3	Identification Strategy . . . . .	34
1.3.1	Empirical specification . . . . .	34
1.3.2	Discussion of identifying assumptions . . . . .	37
1.4	Data and Summary Statistics . . . . .	39
1.4.1	Data sources . . . . .	39
1.4.2	Sample selection . . . . .	40
1.4.3	Main variables . . . . .	40
1.4.4	Summary statistics . . . . .	43
1.5	Main Results . . . . .	46
1.5.1	Gender funding gap within sectors . . . . .	46
1.5.2	The effect of gender stereotypes . . . . .	48
1.5.3	Future corporate outcomes . . . . .	49
1.5.4	Mechanism: Higher bar for minorities or selected pool of entrepreneurs? . . . . .	54
1.6	Robustness Tests . . . . .	55
1.6.1	Human capital . . . . .	55

1.6.2	Team and family . . . . .	56
1.6.3	Motivations and optimism . . . . .	57
1.6.4	The effect of age stereotypes . . . . .	58
1.7	Further Alternative Explanations . . . . .	59
1.7.1	Out of sample replication . . . . .	59
1.7.2	Investor gender, homophily and future corporate performance	61
1.8	Conclusion . . . . .	62
1.9	Figures and Tables . . . . .	65
1.10	Description of variables . . . . .	84
1.11	Appendix . . . . .	86
1.11.1	Additional Tables . . . . .	86
1.11.2	Scrapping . . . . .	105
<b>2</b>	<b>Build or Buy? Human Capital and Corporate Diversification</b>	<b>107</b>
2.1	Introduction . . . . .	109
2.2	Theoretical framework: “build” or “buy”?	116
2.2.1	Basic framework . . . . .	116
2.2.2	Testable predictions: build or buy?	118
2.2.3	Micro-foundations of the labor cost . . . . .	118
2.3	Empirical strategy . . . . .	121
2.3.1	Occupation-specific human capital . . . . .	121
2.3.2	Firm’s internal human capital . . . . .	122
2.3.3	Empirical model . . . . .	124
2.4	Data and summary statistics . . . . .	125
2.4.1	Data sources . . . . .	125
2.4.2	Main variables . . . . .	128
2.4.3	Summary statistics . . . . .	130
2.5	Human capital and corporate diversification . . . . .	131
2.5.1	Main results . . . . .	131
2.5.2	The role of size and financial constraints . . . . .	133
2.5.3	The role of sectoral and geographical distances . . . . .	134
2.6	Robustness checks and alternative mechanisms . . . . .	135
2.6.1	Alternative measures of human capital . . . . .	135

2.6.2	Does selection into diversification drive our results? . . . . .	136
2.6.3	The role of scale and physical capital . . . . .	137
2.7	Evidence of labor adjustment costs . . . . .	138
2.7.1	Within-firm human capital and diversification choice . . . . .	138
2.7.2	Internal human capital and workforce adjustment . . . . .	139
2.8	The role of labor market frictions . . . . .	140
2.8.1	The effects of local labor market tightness for key occupations	140
2.8.2	Local labor market tightness and the value of building . . . . .	141
2.8.3	Build or buy and firing costs . . . . .	141
2.9	Conclusion . . . . .	142
2.10	Figures and Tables . . . . .	143
2.11	Appendix . . . . .	151
2.11.1	Theoretical Appendix . . . . .	161
2.11.2	Additional Tables and Figures . . . . .	165
2.12	Description of Variables . . . . .	173
<b>3</b>	<b>Are Investors Aware of Ownership Connections?</b>	<b>176</b>
3.1	Introduction . . . . .	178
3.2	Sample selection . . . . .	185
3.2.1	Ownership links . . . . .	185
3.2.2	Earnings surprises . . . . .	186
3.2.3	Investors' reactions . . . . .	187
3.3	Description of the sample . . . . .	188
3.3.1	Geographic breakdown . . . . .	188
3.3.2	Descriptive statistics . . . . .	189
3.4	Empirical results . . . . .	189
3.4.1	Investor Reactions to Earnings Announcements . . . . .	189
3.4.2	Relative announcement timing and information incorporation	191
3.4.3	Channels of investor unawareness . . . . .	196
3.4.4	Corporate complexity . . . . .	197
3.4.5	Investor sophistication . . . . .	197
3.5	Robustness Checks . . . . .	198
3.5.1	Endogenous strategic announcement timing . . . . .	198

3.5.2	Tunneling and parents' expropriation behavior . . . . .	199
3.5.3	Internal capital markets . . . . .	200
3.5.4	Confounding events . . . . .	202
3.5.5	Analysts' updating their forecasts after the first announcement	203
3.6	Conclusion . . . . .	203
3.7	Tables . . . . .	204

# Chapter 1

## Mind the Gap: Gender Stereotypes and Entrepreneur Financing

\*

---

\*This chapter benefited from the comments and discussions with Paul Beaumont, Sylvain Catherine, Yixin Chen, Alberta di Giuli, Marie Ekeland, Daniel Ferreira, Zsuzsanna Fluck, Jasmin Gider, Edith Ginglinger, Paul Goldsmith-Pinkham, Denis Gromb, Johan Hombert, Jessica Jeffers, Victor Lyonnet, Thorsten Martin, Adrien Matray, David Matsa, Debarshi Nandy, Paige Ouimet, Thomas Philippon, Manju Puri, Luc Renneboog, David Robinson, Oliver Spalt, Geoffrey Tate and Tracy Wang, as well as seminar participants at BI Oslo, Università Bocconi, the Colorado Finance Summit, University of Colorado Boulder, Dartmouth College, EFA 2019, the 3rd French corporate governance workshop, FIRS 2018, the 4th HEC PhD workshop, HEC Paris, Indiana University, IFN, INSEAD, University of Maryland, NFA 2018, University of New South Wales, University of North Carolina, Northwestern University, Ohio State University, the Paris December Finance Meeting, Université Paris-Dauphine, RBFC 2018, University of Rochester, University of Southern California, Stockholm School of Economics, Tilburg University, University of Toronto, Vanderbilt University, University of Washington and at the WFA 2019. This chapter received the Cubist Systematic Strategies Ph.D. Award for Outstanding Research at the WFA 2019.

## **Abstract**

Using administrative data on the population of start-ups in France and their financing sources, I provide evidence consistent with the existence of stereotypes among equity investors. First, I find that female-founded start-ups are 19-27% less likely to raise external equity including venture capital. However, in female-dominated sectors, female-founded start-ups are no longer at a disadvantage. They are equally to 8% more likely to be backed with equity relative to male-founded start-ups in those sectors and to female-founded start-ups in male-dominated sectors. My empirical design ensures that the observed gender funding gaps are not driven by the composition of founding teams or by differences across individuals regarding human capital, ex ante motivations and optimism. Second, consistent with the idea that the bar is set higher for minorities, I find that conditionally on being backed with equity, female (male) entrepreneurs perform better in male (female)-dominated sectors relative to male (female) entrepreneurs. The evidence is consistent with a model in which investors have context-dependent stereotypes.

## 1.1. Introduction

Is it worth being different? The large literature on discrimination against gender and racial minorities suggests it is not. For example, within symphony orchestras, female musicians are less likely to be hired (Goldin and Rouse, 2000). In the US, “Lakisha” and “Jamal” are less likely to be invited for an interview than “Emily” and “Greg” (Bertrand and Mullainathan, 2004). In the mutual fund industry, managers with foreign-sounding names and female managers receive fewer fund flows and are less likely to be promoted (Kumar, Niessen-Ruenzi and Spalt, 2015; Niessen-Ruenzi and Ruenzi, 2018; Barber, Scherbina and Schlusche, 2017). At S&P 500 firms, women make up 19% of board members and merely 5% of CEOs (Adams and Ferreira, 2009). Within a male-dominated academic field, such as economics, 35% of new PhDs are female, and 12% hold a full professorship (McElroy, 2016; Sarsons, 2017; Chari and Goldsmith-Pinkham, 2017). Finally, in high growth entrepreneurship, while female entrepreneurs represent approximately 30% of the population of start-up founders across time and countries, 10-15% of them succeed in receiving private equity (PE) and venture capital (VC) financing (Gompers and Wang, 2017*b*; Kauffman, 2017; MIWE, 2018). In this paper, I ask whether female entrepreneurs are systematically at a disadvantage in raising capital, and whether it is still the case in environments where they constitute the dominant group. The answers have implications for determining the optimal regulatory response, if any, and more broadly, for understanding how investors’ beliefs affect the development of young firms.

Many explanations have been proposed to rationalize the gender gap, including differences in human capital accumulation, risk attitudes and preferences.<sup>2</sup> These differences imply that women are not drawn into entrepreneurship at all or that they are, but with different motivations and in different industries. Another body of the literature focuses on discrimination and suggests that the gender gap may be due to a lower propensity for investors to fund female entrepreneurs seeking capital. This view stems from the fact that over 90% of venture capitalists (VCs) are men, resulting in difficulties in selecting and advising female entrepreneurs (Gompers et al., 2014; Ewens and Townsend, 2017; Raina, 2017). Nevertheless, it is also

---

<sup>2</sup>See, for instance, Niederle and Vesterlund (2007), Sapienza, Zingales and Maestripieri (2009), Ors, Palomino and Peyrache (2013), Cook et al. (2018), and Bertrand (2011) for a review of the literature.

possible that some investors may be biased against women.<sup>3</sup> A third view related to stereotypes posits that investors underestimate the abilities of entrepreneurs when they belong to the minority (Bordalo et al., 2016). In this paper, I find that female start-up founders are not systematically at a disadvantage in raising capital from external equity investors. They are in sectors in which according to context-dependent stereotypes male entrepreneurs are perceived to do better than female entrepreneurs.

A key challenge for my study is that entrepreneurs' abilities cannot be directly observed. We do not know whether start-ups that did not raise capital had their applications rejected because they were objectively lower-quality projects than those that were funded, or for other reasons. The profile of firms that could use VC but do not, could provide a useful counterfactual to understand what makes a good candidate from investors' point of view. In addition, the underrepresentation of female entrepreneurs among successfully funded entrepreneurs does not necessarily point toward a differential treatment of women by investors, only the *disproportion* between funded entrepreneurs and their representativeness in the population of start-ups does. However, traditional datasets only provide information about firms that have successfully raised capital in public or private equity markets.

In this paper, I take advantage of administrative data from France. The dataset combines a large-scale survey of entrepreneurs with corporate tax files from 2002 to 2016. Every four years, a new cohort of randomly selected entrepreneurs that represents approximately 25% of the population of new firms founded within a year is required to take part in the survey. The first advantage of using administrative data is that the dataset is not subject to the selection biases commonly encountered in the empirical entrepreneurship literature. Second, because I follow full cohorts of entrepreneurs, I can compare the proportion of successfully funded entrepreneurs from a certain gender group to the frequency of this group in the sector. Third,

---

<sup>3</sup>In the summer of 2017, several cases of discrimination against women in technology companies (e.g., Uber, Google) and VC firms (e.g., Kleiner Perkins Caufield & Byers, 500 Startups) highlighted the treatment of women in Silicon Valley (source: <https://goo.gl/VmLJNq><https://goo.gl/VmLJNq>). Other anecdotal evidence involve, for instance, the financier John Doerr who summed up his philosophy as follow: "Invest in white male nerds who've dropped out of Harvard or Stanford", or the Witchsy cofounders who created a fake male cofounder named "Keith Mann" to reach VCs via email and received an unprecedented number of replies.

for each firm, the dataset contains detailed project characteristics, including the activity and financing sources available to the start-up. It also includes a large range of founders' biographical characteristics and personality traits. Specifically, entrepreneurs are asked *ex ante* about their motivations for founding a start-up and their ambitions for the new venture. This qualitative information is likely to matter when investors select start-ups to finance. Fourth, because the corporate tax files include balance sheets, income statements, and employment composition of every firm in France every year, I can characterize and quantify differences in growth and performance between minority-led firms and non-minority-led firms in the early part of their life cycle to shed light on some of the outstanding questions on the role of entrepreneurs' abilities in new firm creation.

My findings are broadly consistent with the view on stereotyping. Although female-founded start-ups are *on average* 19-26% less likely to be financed by external equity investors, I find that this gap no longer exists in female-dominated sectors.<sup>4</sup> Female entrepreneurs in female-dominated sectors are equally to 8% more likely to raise capital relative to male entrepreneurs in those sectors, and significantly more likely to raise capital relative to female entrepreneurs in male-dominated sectors. This finding indicates that both female and male entrepreneurs benefit from operating in a sector in which they fit the representative gender.

To interpret the evidence, I propose a framework based on Bordalo et al. (2016). The model generates empirical predictions for when investors are rational, biased against a gender, and have stereotypes. In the model, entrepreneurs of different genders (male or female) and different ability types (high or low) are distributed across industries. Based on the distribution of each gender by industry, I identify the most representative gender and classify industries as male-dominated (e.g., engineering) or female-dominated (e.g., hairdressing) (Gennaioli and Shleifer, 2010). Investors are biased against a gender if they systematically underfund this group regardless of the context and the entrepreneurs' abilities (taste-based discrimination, Becker, 1957). Investors are rational when they select entrepreneurs according to the true

---

<sup>4</sup>Male- and female-dominated sectors are classified according to the gender distribution of entrepreneurs by sector which identifies the most representative gender for each sector. The baseline measure defines a sector as female-dominated if it comprises more than 50% of females among its population of entrepreneurs. Those sectors represent 15% of the sectoral classification. I also provide alternative measures based on the percentage in the populations of female CEOs, female business owners and female business owners at newly created firms.

average abilities of their gender group in the industry (statistical discrimination, Phelps et al., 1972; Arrow, 1973). Lastly, investors have *context-dependent stereotypes* when their investment decisions favor entrepreneurs when their gender is the most representative of an industry. Therefore, the average abilities of entrepreneurs who belong to the representative gender group are overestimated, and underestimated when they belong to the minority group. As a result, male entrepreneurs have a higher probability of raising capital in male-dominated industries than in female-dominated industries, and female entrepreneurs are more likely to be funded in female-dominated industries. This view is consistent with the pattern I find in the data. The empirical evidence suggests that investors are not systematically biased against female entrepreneurs and act according to context-dependent stereotypes as opposed to fixed preferences toward a gender.

Although female entrepreneurs are not systematically at a disadvantage and are more likely to raise capital in female-dominated sectors, it could still be the case that investors have rational expectations about gender abilities within and across sectors. Women could simply be better at female activities and men better at male activities. To determine whether investors' beliefs about gender are rational or not, I design an "outcome test" in the spirit of Becker (1993).<sup>5</sup> I consider the effect of receiving external equity on future corporate performance. The approach consists of comparing the growth and performance of successfully funded start-ups up to five years after receiving external equity. My findings suggest female-led start-ups founded in male-dominated sectors grow faster than male-led start-ups in these sectors, and perform better relative to female-led start-ups in female-dominated sectors. Specifically, I find that successfully funded start-ups run by a female entrepreneur in a male-dominated sector hire more employees, sell more abroad, have a productive use of assets and are more likely to exit by IPOs relative to those incorporated in female-dominated sectors.<sup>6</sup> Interestingly, I find similar results for male entrepreneurs when they constitute the minority group, i.e., in female-dominated sectors. The evidence suggests that the bar to be backed with equity for the marginal entrepreneur

---

<sup>5</sup>See Arnold, Dobbie and Yang (2018) and Dobbie et al. (2018) for an application to racial bias in the bail market and in the consumer lending market.

<sup>6</sup>However, the results show that entrepreneurs who belong to the minority group have systematically worse returns on assets (ROA). The ROA is traditionally used to measure performance of established firms and may be unstable for young firms that are still growing (Puri and Zarutskie, 2012).

who belongs to a minority group is higher than that for the one who belongs to the dominant group. This finding is consistent with the empirical predictions of context-dependent stereotypes.

An alternative interpretation of the better-observed performance is related to the quality of the pool of entrepreneurs. Entrepreneurs may self-select in industries in which they fit the expected gender because they may derive extra utility from behaving according to the social prescriptions (Akerlof and Kranton, 2000; Jouini, Karehnke and Napp, 2018). As a result, the pool of entrepreneurs from the dominant gender group would be of worse quality than the pool of minority entrepreneurs (Kumar, 2010). I find that serial female entrepreneurs as well as female entrepreneurs who start with a new idea of product are more likely to opt for a male-dominated sector as opposed to a female-dominated sector, whereas those who start to enjoy the private benefits of being their own boss are more likely to start in a female-dominated sector. However, when I focus on the *selected* subsample of successfully funded entrepreneurs, these differences disappear (Adams and Funk, 2012; Adams and Ragunathan, 2017).<sup>7</sup> This finding suggests that minority entrepreneurs who pass the selection bar by equity investors are not necessarily different on observables from those, also selected, who belong to the dominant group of a sector.

An alternative explanation for the better performance of the minority group could be screening discrimination (Cornell and Welch, 1996). According to this view female fund managers are better at selecting and advising female entrepreneurs.<sup>8</sup> Using extracts of PE and VC deals from the commercial database Thomson VentureXpert linked to the matched employer-employee dataset, I identify the gender of fund managers and test for this hypothesis. I do not find that female fund managers explain the better performance of female entrepreneurs. One caveat is that a few female general partners manage PE and VC funds, offering too little variation on the supply side to identify a significant relationship.<sup>9</sup>

---

<sup>7</sup>Kumar (2010) finds that female financial analysts perform better than their male counterparts, suggesting that women who self-select into male-dominated occupations are not representative of the population. Adams and Funk (2012) and Adams and Ragunathan (2017) argue that women who sit in boards and reach top corporate positions are not necessarily different from men in those positions.

<sup>8</sup>This explanation is similar to what Jannati et al. (2016) identify as in-group bias and what Gompers et al. (2014) and Gompers and Wang (2017b) identify as homophily.

<sup>9</sup>Approximately 9 % of PE and VC investment firms are run by a female fund manager in my sample. This figure is consistent with what Gompers et al. (2014) and Gompers and Wang (2017b)

Regarding alternative funding sources available to start-ups, I do not find any differences in fundraising success by gender across sectors. Male and female entrepreneurs are equally likely to raise bank debt and receive equity grants supported by governmental programs in both male-dominated and female-dominated sectors. However, the results show that female entrepreneurs rely more on personal debt and non-bank debt. This mixed finding suggests that entrepreneurs from the minority group only partially manage to shift their demand toward alternative external financing sources.<sup>10</sup> In addition, to try to mitigate concerns relative to the fact that female entrepreneurs may be less likely to get external equity financing because they are less likely to demand in the first place, I run three robustness tests. First, I condition the sample on entrepreneurs who declare to have the ambition to grow the start-up, as opposed to stay small, or entrepreneurs who succeeded at getting a bank loan, a personal or a grant, and I still find that female entrepreneurs are less likely to contract with equity investors. Second, I check whether personal resources invested by female entrepreneurs at creation, bank debt and other external financing sources represent different percentages of the start-up's capital structure, and they do not. Third, in an additional robustness test, I control for the amount of personal resources invested at creation by the founder and show that the main findings hold.<sup>11</sup>

Besides, the difference in fundraising success by gender across sectors is robust to an array of start-ups' characteristics and founders' personal traits. In particular,

---

find in their sample. In a robustness tests, I define a dummy variable that takes the value one if at least one women sit in the board of the VC firm, measured as by being in the top five most compensated employees of the VC firm, and find similar results.

<sup>10</sup>Prior studies focusing on bank loans find that female entrepreneurs pay more for credit than do male entrepreneurs (Bellucci, Borisov and Zazzaro, 2010; Alesina, Lotti and Mistrulli, 2013). The specific features of external equity financing relative to bank loans regarding selection and monitoring efforts can explain why equity investors tend to pay more attention to the entrepreneurs' profiles, especially at early stage, relative to other types of fund providers (Casamatta, 2003; Winton and Yerramilli, 2008; Bottazzi, Da Rin and Hellmann, 2016). First, banks typically lend to a wide variety of firms, whereas start-ups with VC tend to have very risky and positively skewed return distributions with a high probability of negative returns and a small probability of extremely high returns. Second, the monitoring process of banks is typically far less intensive than that of VCs. Banks monitor to minimize negative outcomes and identify worsening collateral quality, whereas VCs monitor more intensively and have extensive control rights, such as board seats and voting rights in the start-ups (Kaplan and Stromberg, 2001; Hellmann and Puri, 2002). Third, VCs impose liquidity restrictions on their limited partners, who in turn demand higher returns from their investment.

<sup>11</sup>Information about amounts invested at creation and percentages of capital structure are available only for some cohorts.

differences in education, past industry experience, prior entrepreneurial experience, motivations, optimism and initial start-up size do not fully explain the observed differences in funding outcomes between minority and majority entrepreneurs. I also consider the influence of starting as a team and of being married (Barber and Odean, 2001). I find that female-led teams are even more likely to be discriminated than male-led teams in male-dominated sectors, but they are also even more likely to balance that disadvantage in female-dominated sectors. Furthermore, I find that the founder's gender no longer explains fundraising success when the start-up is founded with the spouse, suggesting that investors value mixed founding team. In addition, using the extracts from VentureXpert linked with corporate tax files, I replicate the main results out-of-sample to address potential concerns about the quality of self-reported data in surveys. I also take advantage of additional information about investors available in VentureXpert to confirm that PE, VC are subject to stereotypical thinking, as opposed to angel investors. Finally, I find that equity investors have stereotypes not only about gender but also about age (Coffman, Exley and Niederle, 2018). I find that entrepreneurs 50 years old or older who operate in young sectors are less likely to raise capital than younger entrepreneurs in those sectors.<sup>12</sup>

There is surprisingly little systematic evidence about the gender gap in financing entrepreneurs, given the public interest in and regulatory concerns about this topic. The few existing studies on gender disparities in high growth entrepreneurship focus on homophily (Gompers and Wang, 2017 $a,b$ ; Raina, 2017; Ewens and Townsend, 2017; Howell and Nanda, 2019). In particular, Ewens and Townsend (2017) and Raina (2017) find that female entrepreneurs are less likely to be targeted by angel investors and perform worse conditionally on being VC-backed, respectively, but the effects disappear when female entrepreneurs are targeted and advised by female investors. In my study, I find evidence of investors' behaviors consistent with context-dependent stereotypes. Gender minorities are less likely to raise capital, but conditional on being backed with equity they perform better.<sup>13</sup>

Taken together, my findings suggest that the *average* investor misses valuable

---

<sup>12</sup>I define young sectors as sectors in which the median CEO age is below 40 years old.

<sup>13</sup>In a randomized control trial conducted on VC and angel investors in 2018, Gornall and Strebulaev (2019) find that female start-up founders receive more feedback of interest in the first step of the application process than similar male founders.

investment opportunities by overlooking minority entrepreneurs. The evidence has important implications from the perspective of entrepreneurs, the VC industry, and the economy in general. First, entrepreneurs' access to external equity financing can make the difference between success and failure, given the advantage of these equity investors in advising start-ups and creating value (e.g., Hellmann and Puri, 2000, 2002; Kaplan and Strömberg, 2003; Kerr, Lerner and Schoar, 2011). Second, not financing the potential success of high-growth oriented entrepreneurs from minorities means that some VCs are deteriorating potentially better performance and are wasting the resources invested by their limited partners (e.g., Gompers and Lerner, 1999; Kaplan and Schoar, 2005). Third, failing to finance entrepreneurs from a minority may ultimately result in missed growth and missed job creation in the economy (e.g., Haltiwanger, Jarmin and Miranda, 2013; Gennaioli et al., 2013; Hsieh et al., 2013).

This study is also related to the economic literature that investigates causes of the gender gap and a more recent stream of literature in finance that studies its effects on various financial and corporate outcomes. More specifically, to highlight the effects of context-dependent stereotypes in the financing of entrepreneurs, I closely follow hypotheses developed in experimental studies and methodologies of existing field studies on the topic. In the lab, Reuben, Sapienza and Zingales (2014) show that stereotypes work against women in math-related tasks, and Coffman, Exley and Niederle (2018) find that employers prefer to hire male over female workers for a male-typed task not because of preferences for gender but because of beliefs. In the field, Arnold, Dobbie and Yang (2018) find evidence consistent with stereotypes in the bail market, Bohren, Imas and Rosenberg (2017) in a math internet forum, and Egan, Matvos and Seru (2017) in the financial advisory industry.

Finally, this paper contributes to a burgeoning literature on behavioral entrepreneurship that has mainly focused on entrepreneurs' personal traits, risk aversion, and overconfidence levels to explain entrepreneurial entry and financial decisions at young firms (Moskowitz and Vissing-Jorgensen, 2002; Landier and Thesmar, 2008; Hurst and Pugsley, 2011; Puri and Robinson, 2013; Hvide and Panos, 2014; Levine and Rubinstein, 2017).<sup>14</sup> My analysis extends this literature by document-

---

<sup>14</sup>See Kerr, Kerr and Xu (2017) for a review of the literature.

ing that high-growth oriented and optimistic entrepreneurs are more likely to raise capital and that external equity investors are also subject to biased beliefs.

## 1.2. Theoretical Framework

In this section, I develop a stylized framework that derives empirical predictions to identify the underlying factors driving the observed investor discrimination behaviors. The model builds on Bordalo et al. (2016), adapts it to the special case of gender discrimination and incorporates alternative explanations of discrimination (Bohren, Imas and Rosenberg, 2017). The framework consists of a financier who learns about an entrepreneur's ability from her gender and industry of incorporation and then uses this information to decide whether to finance her.<sup>15</sup>

### 1.2.1. Set-up

**Entrepreneurs.** Consider an entrepreneur who has a deterministic gender  $g \in \{M, F\}$ , and who started a company in industry  $i \in \{I_M, I_F\}$ . A proportion  $\omega$  of entrepreneurs choose to start in  $I_M$ , so  $\omega$  represents the size of industry  $I_M$ , and  $1 - \omega$  represents the size of industry  $I_F$ . Within industry  $i$ , there is a frequency  $\pi_{g,I} = Pr(G|I)$  that an entrepreneur is of gender  $g$ . Because F and M are complementary types in the population ( $-G \subseteq \Omega - G$ ), the frequency of one gender can be expressed as a function of the other.  $\pi_i$  and  $1 - \pi_i$  denote the frequency of female and male entrepreneurs in industry  $i$ , respectively. I define industries  $I_M$  and  $I_F$  such that  $Pr(F|I_F) > Pr(F|I_M)$ . In addition, an entrepreneur is characterized by an unobservable ability type: she can be a high-ability type individual ( $H$ ) or a low-ability type individual ( $L$ ). Within industry  $i$ , there exists an unobservable proportion  $Pr(H|G, I)$  of entrepreneurs of gender  $g$  who are high-ability type individuals.

**Financiers.** A set of financiers evaluates the entrepreneurs' abilities. For simplicity, I assume there is one financier or a homogeneous set of financiers who select entrepreneurs to finance. Ideally, a rational financier (or one who believes himself

---

<sup>15</sup>In Bordalo et al. (2016), the type is the entrepreneur's gender  $g$  and the population subgroup  $g$  is the industry  $i$  in which a start-up is incorporated.

to be so) wants to finance only high-ability type entrepreneurs, so the probability that an entrepreneur of gender  $g$  incorporated in industry  $i$  is successful at raising external financing is  $Pr(S|G, I, H)$ . However, distributions of entrepreneurs' ability types are not observable, such that the financier may make mistakes and finance a proportion  $Pr(S|G, I, L)$  of low-ability type entrepreneurs at the expense of entrepreneurs of high ability who belong to the other gender group (budget constraint). Therefore, an entrepreneur of gender  $g$  in industry  $i$ 's probability of raising external financing depends on her perceived ability  $\{\widehat{H}, \widehat{L}\}$ , which could be different from her true ability  $\{H, L\}$ . Figure 1.1 presents the decision problem considering two entrepreneur gender types ( $M$  and  $F$ ) and two entrepreneur ability types ( $H$  and  $L$ ). Entrepreneurs are split into two industries ( $I_M$  and  $I_F$ ), in which male and female entrepreneurs, respectively, represent a larger proportion of entrepreneurs.

[Insert figure 1.1 here]

[Fundraising success] An entrepreneur of gender  $g$  in industry  $i$  whose perceived ability is  $Pr(\widehat{H}|G, I)$  has the following probability of being successfully funded:

$$Pr(S|G, I, \widehat{H}) = Pr(\widehat{H}|G, I) \times Pr(G|I) \times Pr(I)$$

where  $Pr(G|I)$  represents the frequency of gender  $g$  in industry  $i$  and  $Pr(I)$  represents the proportion of entrepreneurs incorporated in  $i$ .

### 1.2.2. Discrimination and funding error

Gender discrimination occurs when a male and a female entrepreneur with the same perceived abilities receive different financing outcomes. Discrimination can also be expressed as the difference between male and female entrepreneurs' financing outcomes in industry  $I$ . [Discrimination] Within-industry discrimination is denoted as follows:

$$D(I) \equiv Pr(S|M, I, \widehat{H}) - Pr(S|F, I, \widehat{H})$$

where  $Pr(S|M, I, \widehat{H})$  and  $Pr(S|F, I, \widehat{H})$  are respectively the proportions of male and female entrepreneurs with high perceived ability who raise capital in industry  $I$ .

There is no discrimination if male and female entrepreneurs are equally likely to raise capital given the frequency of each gender within an industry, such that  $I$ ,  $Pr(S|F, I, \hat{H}) = Pr(S|M, I, \hat{H}) \cdot \left(\frac{\pi_I}{1-\pi_I}\right)$ . The term  $\frac{\pi_I}{1-\pi_I}$  accounts for differences in male and female entrepreneurs' participation within a sector. Discrimination occurs when male and female entrepreneurs, who are perceived to be equally able, experience different funding outcomes. For instance, female entrepreneurs are less likely to raise capital than male entrepreneurs, formally  $Pr(S|F, I, \hat{H}) < Pr(S|M, I, \hat{H}) \cdot \left(\frac{\pi_I}{1-\pi_I}\right)$ , or male entrepreneurs are less likely to raise capital, formally  $Pr(S|F, I, \hat{H}) > Pr(S|M, I, \hat{H}) \cdot \left(\frac{\pi_I}{1-\pi_I}\right)$ , whereas entrepreneurs from both groups are *on average* perceived as equally able,  $Pr(\hat{H}|M, I) = Pr(\hat{H}|F, I)$ .

The corollary of discrimination is funding error. Funding error corresponds to the proportion of low-ability type entrepreneurs who successfully raise capital. Funding error can also arise from financiers who make a mistake by categorizing low-ability type entrepreneurs as high-ability type entrepreneurs. In this case, funding error is defined as the difference between successfully funded entrepreneurs perceived as high-ability types and those who truly are high-ability types. [Funding error] Within-industry funding error is denoted as follows:

$$E(I) = Pr(S|G, I, L) \equiv |Pr(S|G, I, \hat{H}) - Pr(S|G, I, H)|$$

where  $Pr(S|G, I, L)$  is the probability that a low-ability entrepreneur of gender  $g$  raises capital in industry  $I$ ,  $Pr(S|G, I, \hat{H})$  is the probability that an entrepreneur perceived as a high-ability type entrepreneur raises capital, and  $Pr(S|G, I, H)$  is the probability that an entrepreneur of high-ability type raises capital.

### 1.2.3. Taste-based discrimination

Taste-based discrimination is rooted in preferences for a gender. Investors are biased *toward* a gender if they consistently favor entrepreneurs of that gender. In contrast, they are biased *against* a gender, if there is a constant distaste associated with that gender. Taste-based discrimination against female entrepreneurs corresponds to the case in which investors have a constant preference for male entrepreneurs over female entrepreneurs ( $C_F > 0$ ). Female entrepreneurs' probability of raising external

financing is systematically lower than that of male entrepreneurs regardless of the context and even if abilities are perceived as equivalent  $Pr(\widehat{H}|F, I) = Pr(\widehat{H}|M, I)$ . Taste-based discrimination against female entrepreneurs leads to  $Pr(S|F, I, \widehat{H}) < Pr(S|M, I, \widehat{H}) \cdot \left(\frac{\pi_I}{1-\pi_I}\right)$  for all industries.

[Taste-based discrimination] If male and female entrepreneurs' abilities are perceived to be equivalent within and across industries, formally  $Pr(\widehat{H}|F, I) = Pr(\widehat{H}|M, I) \forall I$ , then, all else being equal, taste-based discrimination against female entrepreneurs exists if:

$$Pr(S|F, I, \widehat{H}) < Pr(S|M, I, \widehat{H}) \cdot \left(\frac{\pi_I}{1-\pi_I}\right), \forall I \quad (1.1)$$

Under the same conditions, taste-based discrimination against male entrepreneurs exists if:

$$Pr(S|F, I, \widehat{H}) > Pr(S|M, I, \widehat{H}) \cdot \left(\frac{\pi_I}{1-\pi_I}\right), \forall I \quad (1.2)$$

where  $\pi_I$  denotes the frequency of female entrepreneurs within industry  $I$ .

In the presence of taste-based discrimination, the aggregate funding error across industries is positive:  $E = \sum_i E(I) = \sum_i Pr(S|G, I, L) > 0$ . Because of funding errors, the average ability of the gender that is systematically overfunded is lower than the average ability of the group that is systematically underfunded. Under the assumption that entrepreneurs' abilities are constant over time, funding errors imply that the future corporate performance of the overfunded group will systematically underperform those of the group that is underfunded.

#### 1.2.4. Statistical discrimination

Statistical discrimination is rooted in rational beliefs. Investors finance entrepreneurs with respect to the perceived abilities of their gender group and assume that these abilities are correctly assessed. The perceived distribution of entrepreneurs' abilities by gender coincide with their true abilities. Therefore, in industries in which investors perceive female entrepreneurs to have higher abilities, female entrepreneurs are more likely to be funded; likewise, in industries in which investors perceive male entrepreneurs to have higher abilities, male entrepreneurs are more likely to raise capital.

[Statistical-based discrimination] If investors correctly assess entrepreneurs' abil-

ity type,  $Pr(\widehat{H}|M, I) = Pr(H|M, I)$  and  $Pr(\widehat{H}|F, I) = Pr(H|F, I)$ , and if male and female entrepreneurs' abilities are perceived to be equivalent,  $Pr(\widehat{H}|M, I) = Pr(\widehat{H}|F, I)$ , then, all else being equal, the probability of fundraising success for female entrepreneurs is:

$$Pr(S|F, I, \widehat{H}) = Pr(S|M, I, \widehat{H}) \left( \frac{\pi_I}{1 - \pi_I} \right), \text{ in } I \quad (1.3)$$

If in industry  $I_F$ , in which female entrepreneurs' perceived abilities are higher than those of male entrepreneurs,  $Pr(\widehat{H}|F, I_F) > Pr(\widehat{H}|M, I_F)$ , then, all else being equal, the probability of fundraising success for female entrepreneurs is:

$$Pr(S|F, I_F, \widehat{H}) > Pr(S|M, I_F, \widehat{H}) \cdot \left( \frac{\pi_{I_F}}{1 - \pi_{I_F}} \right), \text{ in } I_F \quad (1.4)$$

If in industry  $I_M$ , in which female entrepreneurs' perceived abilities are lower than those of male entrepreneurs,  $Pr(\widehat{H}|F, I_M) < Pr(\widehat{H}|M, I_M)$ , all else being equal, the probability of fundraising success for female entrepreneurs is:

$$Pr(S|F, I_M, \widehat{H}) < Pr(S|M, I_M, \widehat{H}) \cdot \left( \frac{\pi_{I_M}}{1 - \pi_{I_M}} \right), \text{ in } I_M \quad (1.5)$$

where  $\pi_{I_F}$  denotes the frequency of female entrepreneurs in  $I_F$  and  $\pi_{I_M}$  denotes the frequency of female entrepreneurs in  $I_M$ .

Within an industry, taste-based discrimination and statistical-based discrimination yield to the same predictions. Both an exogenous parameter  $C_F > 0$  and beliefs about gender abilities, such as  $Pr(\widehat{H}|M, I) > Pr(\widehat{H}|F, I)$ , would lower female entrepreneurs' fundraising success.<sup>16</sup> I disentangle taste-based discrimination from statistical discrimination by introducing sectoral heterogeneity (proposition 1.2.4). Therefore, I consider two types of industries: male-dominated ( $I_M$ ) and female-dominated ( $I_F$ ). Asymmetric entrepreneur funding outcomes by gender across sectors identify investors' belief-based behaviors, as opposed to preference-based discrimination which predicts that a certain group is consistently underfunded *regardless* of the industry.

---

<sup>16</sup>Note that in both the taste-based discrimination and the statistical discrimination views, investors correctly assess distributions of entrepreneurs' abilities. The difference comes from the fact that investors who rely on preference simply do not use entrepreneurs' abilities when they select entrepreneurs to finance.

In the rational belief-based discrimination view, the *average* funding error by gender within and across sectors is equal to zero.<sup>17</sup> There is no systematic mistake made about the same gender, formally,  $E = E(I_F) + E(I_M) = |Pr(\widehat{H}|M, I_F) - Pr(H|M, I_F)| + |Pr(\widehat{H}|F, I_M) - Pr(H|F, I_M)| = 0$ . Assuming that entrepreneurs have constant abilities over time, successfully funded entrepreneurs who belong to a particular gender group should not display better future performance than those of the other group. Entrepreneurs are financed according to the true ability of their gender group, such as on average minority entrepreneurs are *not* less likely to be funded relative to entrepreneurs from the dominant group.

### 1.2.5. Discrimination with stereotypes

Investors have context-dependent stereotypes if they favor a gender that is representative of the industry. As in statistical discrimination, the financier selects entrepreneurs according to his expectations about the average ability of a gender by industry. However, as in Bordalo et al. (2016), the financier have a distorted view of entrepreneurs abilities by gender across industries.

At firm creation, investors may not have much information about entrepreneurs' abilities and may use entrepreneurs' frequencies by gender ( $Pr(G|I)$ ) to estimate entrepreneurs' abilities. Therefore, the frequency distribution of gender  $g$  is mapped to the ability distribution of entrepreneurs of gender  $g$ ; formally, I assume  $Pr(H|G, I) = Pr(G|I)$  (called "congruity theory" in Eagly and Karau (2002)). Following Gennaioli and Shleifer (2010), the most representative gender  $g$  for industry  $I$  is the one that is most representative of the industry relative to other industries  $-I$ . The representative gender is also the easiest to recall (also called heuristics in Tversky and Kahneman (1983)), e.g., female for the hairdressing industry and male for the engineering industry.

[Representativeness] The representativeness of a gender  $g$  for industry  $I$  given another industry  $-I$  is defined as the likelihood ratio:

$$R(G, I, -I) \equiv \frac{\pi_{g,I}}{\pi_{g,-I}}$$

---

<sup>17</sup>This does not mean that funding errors do not exist at the individual level. Nevertheless, they are not systematically directed toward the same gender as errors are expected to cancel when aggregated.

Gender representativeness captures the fact that a gender is more likely to be overweighted relative to its true frequency if it is unlikely in other industries. Following Bordalo et al. (2016), the financier relies on stereotypes when her beliefs have the following form:

[Distortion] The financier attaches to each gender  $g$  in industry  $I$  a distorted probability:

$$Pr(\widehat{H}^*|G, I) = \frac{\pi_I \cdot h_g\left(\frac{\pi_I}{\pi_{-I}}\right)}{\pi_I \cdot h_g\left(\frac{\pi_I}{\pi_{-I}}\right) + (1 - \pi_I) \cdot h_{-g}\left(\frac{\pi_{-I}}{\pi_I}\right)}$$

where  $\pi_{-I}$  is the frequency of entrepreneurs with gender  $g$  in industry  $I$ , and function  $h_g(\cdot)$  is a symmetric function centered on the representativeness of a gender to an industry; it increases in its own representativeness and decreases in the representativeness of the other gender.

Under this formulation, distorted abilities are modeled as an exaggeration of true gender frequency distributions. If gender  $g$  is objectively more likely within an industry, namely  $Pr(G|I)$  is higher, then the stereotypes imply that the financier overestimates the probability of highly able entrepreneurs who belong to this gender group. As a result, distortions are due exclusively to the fact that one gender is more or less representative of an industry than the other. If all genders are equally representative of an industry, the financier does not distort the true gender distributions of abilities, so he holds rational expectations about genders and  $h(1)$ . If the representativeness of genders differs across industries, stereotypical beliefs outweigh the ability of the most representative gender. Then, following Bordalo et al. (2016), I define ability distributions distorted by context-dependent stereotypes,  $Pr(\widehat{H}^*|G, I)$ , and I compare them to the true distributions of abilities,  $Pr(H|G, I)$ .

[Perceived abilities with stereotypes] If female is the representative gender of industry  $I_F$  and male the representative gender of industry  $I_M$  and assuming that the likelihood ratio  $\frac{\pi_{G, I_F}}{\pi_{G, I_M}}$  is monotonically and strictly increasing in the proportion of a gender  $G = \{M, F\}$ , then for any weighting function  $h_g(\cdot)$ :

$$Pr(\widehat{H}^*|F, I_F) > Pr(H|F, I_F) > Pr(H|F, I_M) > Pr(\widehat{H}^*|F, I_M) \quad (1.6)$$

and

$$Pr(\widehat{H}^*|M, I_F) < Pr(H|M, I_F) < Pr(H|M, I_M) < Pr(\widehat{H}^*|M, I_M) \quad (1.7)$$

Context-dependent stereotypes amplify differences in gender distributions across industries. In particular, the financier overestimates the abilities of female entrepreneurs in female-dominated industries and underestimates their abilities in male-dominated industries. For instance, hairdressing is a female-dominated industry, so the proportion of female hairdressers who are perceived as highly able is higher than the proportion of truly highly able entrepreneurs. In contrast, the proportion of male hairdressers perceived as highly able is lower than the true proportion. The inverse applies to software programming that is a male-dominated industry. The proportion of male programmers perceived as highly able is overestimated, and the proportion of highly able female programmers is underestimated.

[Discrimination with stereotypes] If the perceived abilities of female entrepreneurs in female-dominated industry  $I_F$  is greater than the perceived abilities of female entrepreneurs in male-dominated industry  $I_M$ , formally  $Pr(\widehat{H}^*|F, I_F) > Pr(\widehat{H}^*|F, I_M)$ , then, all else being equal, the probability of fundraising success for female entrepreneurs is:

$$Pr(S|F, I_F, \widehat{H}^*) > Pr(S|F, I_M, \widehat{H}^*) \cdot \left( \frac{\pi_I}{1 - \pi_I} \right) \quad (1.8)$$

If the perceived abilities of male entrepreneurs in male-dominated industry  $I_M$  is greater than the perceived abilities of male entrepreneurs in female-dominated industry  $I_F$ , formally  $Pr(\widehat{H}^*|M, I_M) > Pr(\widehat{H}^*|M, I_F)$ , then, all else being equal, the probability of fundraising success for male entrepreneurs is:

$$Pr(S|M, I_F, \widehat{H}^*) < Pr(S|M, I_M, \widehat{H}^*) \cdot \left( \frac{\pi_I}{1 - \pi_I} \right) \quad (1.9)$$

where  $\pi_{I_F}$  denotes the frequency of female entrepreneurs in  $I_F$  and  $\pi_{I_M}$  denotes the frequency of female entrepreneurs in  $I_M$ . In addition, discrimination with stereotypes exists if  $Pr(\widehat{H}^*|M, I) > Pr(H|M, I)$  and  $Pr(\widehat{H}^*|F, I) > Pr(H|F, I)$ .

Context-dependent stereotypes about gender also yield asymmetric investor funding behaviors. A representative gender is more likely to be funded in those industries in which it is representative as opposed to industries in which it represents the minority group.

As a result, with context-dependent stereotypes, the financier makes systematic funding errors against the minority gender group across industries:  $E = E(I_F) + E(I_M) = Pr(S|F, I_F, L) + Pr(S|M, I_M, L) \equiv |Pr(\widehat{H}^*|M, I_F) - Pr(H|M, I_F)| + |Pr(\widehat{H}^*|F, I_M) - Pr(H|F, I_M)| > 0$ . Assuming that entrepreneurs have constant abilities over time, successfully funded entrepreneurs from the representative group are expected to underperform relative to their performance in industries in which they belong to the minority group. The reason is that a non-zero share of successfully funded entrepreneurs from the representative group is low-ability type entrepreneurs. Empirically, in male-dominated industries, we expect the future performance of successfully funded female entrepreneurs to be greater than that of successfully funded female entrepreneurs in female-dominated industries. The symmetric case applies to successfully funded male entrepreneurs in female-dominated industries.

### 1.2.6. Aggregate effects and policy implications

In this section, I characterize the aggregate effects of stereotypes on the economy. If we consider more than two industries or two industries of different size, formally  $\omega \neq \frac{1}{2}$ , all else being equal, equation 1.2.5 yields the following proposition. [Aggregate effects] If the size of female-dominated  $I_F$  represents less than half of the total economy, formally  $\omega < \frac{1}{2}$ , then the aggregate probability of fundraising success of female entrepreneurs is:

$$Pr(S|F) < Pr(S|M) \tag{1.10}$$

If male-dominated industries account for a larger share of the economy than female-dominated industries, stereotypes favoring male entrepreneurs dominate those favoring female entrepreneurs, and the probability of female entrepreneurs successfully raising capital becomes lower than that of male entrepreneurs. In my framework, this finding is not driven by any form of investor preference but due to the fact

that financiers mistakenly overestimate the abilities of entrepreneurs who belong to the representative group of an industry. In particular, if male- or female-dominated industries were of equal size, the probability of entrepreneurs' fundraising success by gender would be equal, even in the presence of distorted ability distributions by gender across industries.

This framework implies that the increasing participation of female entrepreneurs attenuates the aggregate effects of gender stereotypes. Gender stereotypes can be attenuated by balancing gender representation *within* industries, i.e., more female entrepreneurs in male-dominated industries and more male entrepreneurs in female-dominated industries.

In practice, initiatives favoring the participation of minorities in industries in which they are underrepresented can take the form of communication campaigns and mentoring programs targeting minorities (Meier, Niessen-Ruenzi and Ruenzi, 2017; Del Carpio and Guadalupe, 2018). They can also consist of indirect actions, such as participation quotas in professional tracks directly leading up industries in which minorities are underrepresented. For instance, participation quotas in training programs that supply pools of potential entrepreneurs, e.g., engineering schools, may be useful to meet this objective (Breda and Ly, 2015).

Different policy actions should be carried out if a preference toward male entrepreneurs is identified as the main underlying source of discrimination. In the case where female entrepreneurs systematically fail at raising funds, funding quotas favoring women could balance the alleged constant distaste against female entrepreneurs (i.e.,  $Quota = C_F$ ). In this spirit, professional angel investors associations and foundations have introduced women-only funding programs (among others, e.g., Pipeline Angels, Built by Girls Ventures, Cartier's Women Initiative).<sup>18</sup> Finally, if female entrepreneurs are identified as inherently less able than men at entrepreneurship and if gender equality is of public interest, training programs closing this gender gap in terms of human capital may be introduced.<sup>19</sup>

---

<sup>18</sup>Diversity quotas that aim to directly address gender disparities have been implemented in other contexts. In particular, board of directors gender quotas exist in several European countries (among others, e.g., Matsa and Miller, 2013; Bertrand et al., 2014).

<sup>19</sup>Differences in human capital, and especially in terms of education have been a classical explanation in the literature to rationalize the gender gaps (Bertrand, 2011).

## 1.3. Identification Strategy

### 1.3.1. Empirical specification

The empirical analysis aims to identify for a population of start-ups whether entrepreneurs' gender matters in the allocation of capital. In particular, female entrepreneurs may have a lower probability of fundraising success, first, because they are different from male entrepreneurs (proposition 1.2.4), second, because investors have a preference toward male entrepreneurs (proposition 1.2.3), third, due to stereotyping (proposition 1.2.5). Empirically, I compare male and female entrepreneurs' funding outcomes within and across sectors. The null hypothesis predicts that gender should not matter after controlling for abilities, formally  $Pr(\widehat{S|F, I, \widehat{H}}) = Pr(\widehat{S|M, I, \widehat{H}}) \forall I$ , all else being equal. In contrast, if gender disparities exist after controlling for abilities, this finding would predict taste-based discrimination. The first empirical specification compares entrepreneurs' probabilities of fundraising success within a sector and is given by the following equation:

$$Success_i = \lambda_z + \lambda_{st} + \delta Female_i + \beta' X_i + \epsilon_i \quad (1.11)$$

where  $Success_i$  is a dummy variable that takes the value one if start-up  $i$  incorporated in sector  $s$  and county  $z$  and belonging to cohort-year  $t$  successfully raises capital, and zero otherwise;  $\lambda_z$  and  $\lambda_{st}$  correspond to zip code and sector  $\times$  cohort fixed effects, respectively; and  $X_i$  represents a vector of additional controls. Specifically,  $X_i$  comprises the start-up's incorporation status; the logarithm of total assets; the ratio of tangible assets; and biographical characteristics of entrepreneurs, such as age, French citizenship, education and work experience dummy variables. All variables are defined in Appendix table 1.14. The main independent variable is the dummy  $Female$  which captures the start-up founder's gender. This models controls for fixed characteristics across sectors and locations. In particular, it accounts for the fact that entrepreneurs with specific ability types may cluster in certain industries and geographies, but also for the fact that investors also specialize in specific sectors and select start-ups in their local areas (Sørensen, 2007).

In this specification, the rational view predicts  $\delta = 0$ , assuming no differences

in abilities or that differences in abilities are perfectly accounted for by the controls (proposition 1.2.4). Taste-based discrimination against female entrepreneurs predicts  $\delta < 0$ , under the same conditions (proposition 1.2.3). Note that the context-dependent stereotype view cannot be identified when comparing entrepreneurs within a sector.

To identify stereotypes, I specify a second test that compares entrepreneurs' funding outcomes across sectors. I classify sectors into two categories, female-dominated and male-dominated. Empirically, I identify a female-dominated sector ( $Female.Sector_t$ ) at the 4-digit SIC level if it has more than 50% female-founded start-ups within a cohort-year. The empirical specification that identifies investors' context-dependent stereotypes is given by the following equation:

$$\begin{aligned} Success_i = & \lambda_z + \lambda_s + \lambda_t + \delta_1 Female_i + \delta_2 Female.Sector_t \\ & + \delta_3 Female_i \times Female.Sector_t + \beta' X_i + \gamma' Z_{st} + \epsilon_i \end{aligned} \tag{1.12}$$

The rational view still predicts  $\delta_i = 0$ ,  $\forall, \delta_i$  with  $i \in \{1, 2, 3\}$ , assuming that the controls perfectly account for gender differences in abilities. The taste-based discrimination view against female entrepreneurs predicts that female entrepreneurs are systematically underfunded across sectors, such that  $\delta_1 < 0$  and  $\delta_3 < 0$ . The taste-based view does not give any prediction for  $\delta_2$ , since the negative relationship between the female gender and the likelihood of raising capital is already captured by  $\delta_1$ . The context-dependent stereotype view predicts asymmetric entrepreneur funding outcomes by gender across sectors (proposition 1.2.5). In particular, female entrepreneurs in female-dominated sectors are more likely to raise capital than men in female-dominated sectors, and than women in male-dominated sectors relative to their own representativeness, such that  $\delta_3 > 0$ ,  $\delta_2 < 0$  and  $\delta_1 < 0$ . According to proposition 1.2.6, the sign of the sum of coefficients  $\delta_1 + \delta_2 + \delta_3$  depends on the share of female-dominated sectors in the economy (parameter  $\omega$  in the model). In particular,  $\delta_1 + \delta_2 + \delta_3 < 0$  when female-dominated sectors represent a minority share of the economy, and  $\delta_1 + \delta_2 + \delta_3 > 0$  when female-dominated sectors represent a majority share of the economy. Note that specification 1.12 compares entrepreneurs within the same sector across time and does not account for unobservable time-

varying sectoral characteristics. Thus, I introduce an additional set of time-varying sector control variables  $Z_{st}$ , which include the sector size, the Herfindahl index, and the frequency of female entrepreneurs within a sector. Including the within-sector percentage of female entrepreneurs ensures that specification 1.12 is not picking up a mechanical relationship between the proportion of female entrepreneurs and their likelihood of raising capital.

Finally, it is still possible that conditioning on observables does not perfectly account for differences in individual abilities. Investors may rationally discriminate against female entrepreneurs in male-dominated industries, and against male entrepreneurs in female-dominated industries, if male entrepreneurs have higher *unobservable* abilities at male activities, and if female entrepreneurs have higher abilities at female activities. To test this hypothesis, I design an “outcome test” in the spirit of Becker (1993). The idea is that we should not observe any systematic difference between male and female entrepreneurs’ future performance if they are selected according to their true abilities. Empirically, I use the number of employees, sales over assets, the share of sales realized abroad, and survival measures from one year after creation to the five onwards as measures of start-ups’ performance. I interact the entrepreneur’s gender with the dummy variable *Male.Sector*, as I am interested in how successfully funded female-founded start-ups perform compared to similar male-founded start-ups in male-dominated sectors, and compared to similar female-founded start-ups in female-dominated sectors. Similarly, I test how perform *Male* entrepreneurs in female-dominated sectors.

$$\begin{aligned}
 Outcome_{i,t=\{t+1,t+5\}} &= \lambda_z + \lambda_s + \lambda_t + \delta_1 Female_i + \delta_2 Male.Sector_t \\
 &\quad + \delta_3 Female_i \times Male.Sector_t + \beta' X_{it} + \epsilon_{i,t}
 \end{aligned}
 \tag{1.13}$$

where  $Outcome_{i,t}$  corresponds to the future start-up’s corporate outcome up to five years after creation. The statistical discrimination view predicts no systematic gender differences in future corporate outcomes, so  $\delta_1 = 0$ ,  $\delta_2 = 0$ , and  $\delta_3 = 0$ . Entrepreneurs by group have been have been backed according to their true average abilities. The context-dependent stereotypes view predicts that the successfully funded female-founded start-ups in male-dominated sectors perform marginally bet-

ter than their male counterparts in male-dominated sectors ( $\delta_1 + \delta_2 + \delta_3 > 0$ ), and better than female-founded start-ups in female-dominated sectors ( $\delta_2 + \delta_3 > 0$ ). Finally, an alternative view called “positive discrimination” would predict that minorities are over-funded in environments in which they are underrepresented, such that the marginal entrepreneurs from the minority group should perform marginally worse than majority-group entrepreneurs ( $\delta_2 + \delta_3 < 0$ ).

### 1.3.2. Discussion of identifying assumptions

The empirical analysis aims to estimate entrepreneurs’ probability of fundraising success by gender across industries to identify potential investors’ discrimination behaviors. I compare the probability of fundraising success  $Pr(S|G, I, H)$  given by the framework to the observed probability of fundraising success. According to definition 1.2.1, the true probability of fundraising success is conditioned on perceived abilities ( $Pr(\hat{H}|G, I)$ ) and depends on relative sector sizes ( $\omega$ ) and the participation rates of each gender by industry ( $\pi_I$ ). Empirically, I observe sector sizes (number of entrepreneurs) and the unbiased gender participation by sector (frequency of female entrepreneurs).<sup>20</sup> However, entrepreneurs’ abilities cannot be directly observed. An ideal specification would introduce entrepreneurs’ fixed effects to capture variation in ability at the individual level. Such specification requires the ability to observe the time series of an entrepreneur’s funding outcomes, i.e., serial entrepreneurs, as well as variation in the entrepreneurs’ gender type. Nevertheless, a few cases of serial entrepreneurs occur in my sample, and in the case of new ventures, the entrepreneur’s gender does not usually vary within firm over time. As a result, the empirical analysis builds on assumptions regarding the sources of variation in entrepreneurs’ abilities.

First, one can assume that entrepreneurs’ abilities are industry-specific, implying that entrepreneurs of the high-ability type cluster in a few industries and those of the low-ability type cluster in different ones, formally  $Pr(\hat{H}|I) > Pr(\hat{H}|-I)$ . The within-sector specification (equation 1.11) includes sector  $\times$  cohort-year fixed effects and captures unobservable heterogeneity in ability across sectors as well as time-

---

<sup>20</sup>Observing the unbiased gender distribution by sector requires the use of administrative data based on national firm registries.

varying sector characteristics, such as sector size, product market concentration, and participation rates by gender across sectors.

Second, one can alternatively assume that entrepreneurs' abilities are gender-specific and do not vary across industries. This is, for instance, the case when one assumes that women have lower abilities as entrepreneurs than men, formally  $Pr(\widehat{H}|F) < Pr(\widehat{H}|M) \forall I$ . This hypothesis can be tested by making the following contrarian argument: if the female entrepreneurs' probability of fundraising success is higher than that of men in at least one industry ( $Pr(\widehat{S}|F, I^*, H) > Pr(\widehat{S}|M, I^*, H)$ ), then female entrepreneurs are not systematically less able than men at entrepreneurship.

Third, entrepreneurs' abilities can vary with both gender and industry. This is the case when one assumes that women are better at female activities, and men better at male activities, formally,  $Pr(\widehat{H}|F, I_F) > Pr(\widehat{H}|F, I_M)$  and  $Pr(\widehat{H}|M, I_M) > Pr(\widehat{H}|M, I_F)$ . This argument is consistent with the idea that entrepreneurs rationally self-select into sectors in which they have better abilities, or in which they derive more utility (also called private benefits) by behaving in accordance with the social prescriptions of their gender (Akerlof and Kranton, 2000).

My answer to this argument is threefold: first, I control for a large range of individual characteristics and personality traits arguably correlated with individual abilities. In all models, I control for education, industry expertise and entrepreneurial experience. In addition, entrepreneurs are asked about their ex ante motivation in creating a start-up (desire for independence, opportunity, taste and new ideas). Entrepreneurs are also asked at founding time whether they intend to develop the start-up or become their own boss and stay small (high-growth oriented entrepreneurs). These motives are arguably correlated with entrepreneurial abilities and efforts. Behavioral traits such as overconfidence, as well as family and team composition, may also be related to entrepreneurial abilities. Further robustness tests address these concerns.

However, it is still possible that entrepreneurs who belong to the minority and the majority groups are different on unobservable dimensions. Second, I test whether minority and majority perform differently after receiving financing. If the bar was set at the right place, we should not observe any differences in performance, assuming

that entrepreneurs' abilities are constant over time. Third, entrepreneurs may not be randomly distributed across sectors and self-select within sector. In an additional test, I endogenize the choice for a female- versus a male-dominated sector. I regress this choice on the aforesaid entrepreneurs' personal characteristics interacted with the entrepreneur's gender for both the entire pool of entrepreneurs by sector and the subsample of equity-backed entrepreneurs. This test captures observable differences in entrepreneurs' quality by gender across sectors.

## 1.4. Data and Summary Statistics

### 1.4.1. Data sources

My dataset consists of the merging of two primary data sources available from the French Bureau of Statistics (Insee). The first source is a survey of entrepreneurs administered to cohorts of entrepreneurs who started businesses in 2002, 2006, 2010, and 2014. Tax files are the second source. They provide detailed yearly accounting and employment information at the firm level between 2002 and 2016.

**Entrepreneurs.** The *Système d'Information des Nouvelles Entreprises* (SINE) survey is a large-scale survey of entrepreneurs conducted by the French Bureau of Statistics every four years (see Landier and Thesmar, 2008; Hombert et al., 2017). Questionnaires are sent to approximately 25% of entrepreneurs who started or took over a business in France that year (cohort). The surveyed firms are randomly selected from firm registries.<sup>21</sup> The response rate to SINE surveys is high (approximately 90%) because the tax authorities supervise the sending of questionnaires. For each cohort, I start with 30,000 to 50,000 firms. Three years after their creation/takeover, these firms are re-sent similar questionnaires, but only 65% of the firms in the initial cohort respond. This attrition is explained by failed businesses and by businesses changing locations and not being located by survey managers. Then, five years after business creation/takeover, a last wave of questionnaires is sent, and the average attrition rate is 45%.

**Tax files.** Tax files (*Bénéfices Industriels et Commerciaux* and *Bénéfices Non-*

---

<sup>21</sup>The firm registry contains the universe of registered firms each month in France from 1993 to 2016.

*commerciaux*) augmented by the employer payrolls (*Déclarations Annuelles des Données Sociales*) are available every year and provide balance sheet information, operating income, and employment composition. These files cover all firms subject to the regular corporate tax regime or the simplified corporate tax regime. Small firms with annual sales below €32,600 (€81,500 in retail and wholesale trade) can opt out and choose a special micro-business tax regime (called *micro-entreprise*). Income falling into this category is taxed at the personal level. These firms do not, therefore, appear in the corporate tax files.<sup>22</sup>

### 1.4.2. Sample selection

The sample is the result of a merger of the SINE survey and corporate tax files. The main dataset consists of a repeated cross-section of firms started during the four cohort-years during which the survey is run. Firms are excluded if they opt for a simplified regime and therefore are not present in the corporate tax files. In addition, to study real start-ups, new entrepreneurs who inherited or took over an already existing business are excluded from the sample. In the analysis, I control for the incorporation status of firms (i.e., incorporated firms and sole proprietorships). Thus, I account for the fact that entrepreneurship aggregates different types of activities and individuals, making little distinction between high-growth oriented entrepreneurs and survival entrepreneurs (Schoar, 2010; Hurst and Pugsley, 2011; Levine and Rubinstein, 2017). The limited liability associated with incorporation reduces the potential downside losses to equity holders, thereby increasing the appeal of projects with high expected returns. Because incorporated firms are legal entities separate from their founders, corporations are allowed to own property and to contract independently with financiers and other stakeholders. The incorporation status is important in my study because those start-ups are more likely to seek external finance and investors are therefore more likely to finance those firms.

### 1.4.3. Main variables

**Financing sources.** I identify the start-ups' financing sources using the SINE survey. Entrepreneurs self-report the financing sources they rely on at creation.

---

<sup>22</sup>See Aghion et al. (2017) for more detail about the different tax regimes in France.

The answers are non-exclusive: an entrepreneur can rely on both internal and external resources. Internal financing denotes personal resources invested at creation, whereas external financing sources are split into debt and equity. External debt comprises *Personal loans* granted to the entrepreneur as a person; *Bank loans* granted to the company; and *Non-bank loans* issued by non-financial institutions and public institutions (e.g., zero interest rate loans).

**External equity** encompasses VC, business equity and equity grants. VC and business equity provisions are pooled and studied indistinguishably because they both involve a high degree of target selection and shareholder activism. In addition, a distinction between these two types of external equity is only possible in the 2002 cohort and from the 2010 cohort onward. For external equity financing, I compare self-reported access to external equity investors to PE deals (VC, other PE, CVC and angel investors) reported in the Thomson VentureXpert over the period.<sup>23</sup> I found a high correspondence between the matched firms in the two datasets. *Subsidies* is a very heterogeneous class that mainly includes equity stipends from various public programs.<sup>24</sup> Subsidies in my context are not specifically designed for high-growth oriented entrepreneurs Howell (2017).

**Biographical and human capital information.** Gender, age, and citizenship dummy variables are collected from the SINE survey. The tax authorities send the questionnaires to the business owner, who is in charge of completing the documents. Human capital information is also obtained from the SINE survey. Education information is re-coded, so that cohorts can be compared across time.<sup>25</sup> Education dummy variables include *No degree*, *High school*, *Bachelor's*, *Master's/PhD*, and *Elite engineering school*. Additionally, entrepreneurs are asked about the number of years they worked in the industry before entry into entrepreneurship. I code a dummy *Expert* if the entrepreneur declares at least three years of industry experience. The dummy variable *Serial* indicates whether the entrepreneur has already

---

<sup>23</sup>Target companies and investment firms involved in deals reported in the Thomson VentureXpert database are matched to the universe of French administrative data using a Python web-crawler. See appendix 1.A21 for more details about the procedure.

<sup>24</sup>Examples of public programs that fall into this category: ACCRE, NACRE, PCE, CIR programs, OSEO innovation grants, and AGEFIPH aid.

<sup>25</sup>In particular, a major reform of the higher education system occurred in 2006 that homogenized university diplomas and made them comparable across European countries.

founded a start-up before the one targeted by the questionnaire.<sup>26</sup>

**Motivations.** Entrepreneurs' motivations to start are arguably correlated with unobservable abilities and effort and are important for understanding what drives demand for specific financing sources. The SINE survey asks entrepreneurs about their desire to grow the founded start-up. The possible answers are "to develop the company" and "to create one's own job". The variable *High-growth oriented* entrepreneur is coded accordingly. In a separate question, entrepreneurs are asked about their three main motivations for entering into entrepreneurship. The respondents choose up to three answers from among the following list: *Add earnings* to the household; desire for *Independence*; address unemployment; follow a *Taste* for entrepreneurship and new challenges; take on an *Opportunity*; and explore a *New idea* for a product, service, or market.

**Optimism.** Behavioral effects may also correlate with abilities and fundraising success. I replicate Landier and Thesmar (2008)'s measure of optimism. Optimism is defined as the difference between initial employment expectations and the actual realization in the following year. An entrepreneur is identified as optimistic if she answers "yes" to the question "Do you plan to hire over the next 12 months?" and then she does not do so. Realized employment growth is therefore equal to zero, as is the second part of the following equation:

Optimism at start = 1(Hiring expectation) - 1( $\Delta$ Employment  $\geq 2$  and Firm survives).

**Sector, location, and firm performance.** In the analysis, I control for geographic location at the county level (*département*). France is divided into 101 département. I also control for industrial activity with 4-digit SIC dummy variables. The French equivalent of the SIC classification (*Nomenclature des Activités Françaises*, NAF) consists of 540 sectors at the 4-digit level.<sup>27</sup> Counties and SIC codes are col-

---

<sup>26</sup>Serial entrepreneurs can be either individuals who run several companies at the same time or who restart a new business after having exited at least once in the past.

<sup>27</sup>A major change in the French nomenclature of activities occurred in 2008 (NAF rev.2) and necessitates the following adjustments to preserve the industry panel structure of the data. In the tax files, the SIC codes before 2008 are given in the old classification. I retrieve the industry codes of firms before 2008 from the retropolated firm registries and use the most likely correspondence between the two classifications if not available in the firm registries.

lected from tax files. From the tax files, I also retrieve accounting information to measure firm sales, employment size, total assets, earnings before interest and tax (EBIT) and net income.

**Gender-dominated sectors.** I assess the effect of context-dependence by determining the most representative gender of each sector at the 4-digit SIC level. A sector is flagged as female-dominated if for a given cohort more than 50% of new firms within a sector are female-founded. The baseline measure is based on newly created firms available in the SINE surveys. For robustness, I construct other measures of female-dominated sectors. They are defined based on the within-sector representativeness of female CEOs, female workers, female business owners and new female business owners at newly created firms. The representativeness of female CEOs and female workers is based on the employer payrolls database, whereas the representativeness of female business owners is identified through firm registries. In further robustness tests, I modify the threshold of 50% to 45% and 55% thresholds, I exclude sectors between 45% and 55% of female representation, and I exclude sectors that switch from one gender to another.

#### 1.4.4. Summary statistics

##### Heterogeneity in the gender gap across sectors

Table 1.1 shows that female entrepreneurs found 26% of the 92,446 new firms in my sample. The gender gap is only slowly closing over time starting from 23% in 2002 to 28% of female entrepreneurs in 2014. These figures confirm the existence of a large and persistent gender gap in entrepreneurial participation.

[Insert table 1.1 here]

Appendix table 1.A1 provides frequencies of female entrepreneurs by sector at the 1-digit French sectoral industrial classification (18 sectors). Female-founded startups are concentrated in healthcare industries (62%), educational activities (40%), and service-related sectors (62%). IT and financial services include 18% and 25% female entrepreneurs, respectively. At the 4-digit SIC level, women tend to sort into service-related activities ("hairdressing and other beauty treatment", 77%) and

healthcare sectors ("other human health activities", 70%). By contrast, "forging, pressing, stamping", the manufacture of "bodies for motor vehicles" and the "repair of electrical equipment" include 3/4% of female-founded start-ups.

I identify female-dominated sectors if at least 50% of start-ups within a cohort-year are female-founded. Appendix table 1.A3 shows that 15% of entrepreneurs sort into female-dominated sectors. Female-dominated sectors attract twice as many women as men, whereas male-dominated sectors involve more than three times as many men as women.

### **Are male and female entrepreneurs different?**

Table 1.2 presents the means and differences between male and female entrepreneurs for various biographical characteristics and personality traits. About half of the entrepreneurs are 40 years or older. Female entrepreneurs are on average younger than male entrepreneurs. They are also more likely to be French citizens (93% versus 90% of males). Regarding education, the average female start-up founder is more educated than the average male entrepreneur: 34% of male entrepreneurs and 43% of female entrepreneurs hold a bachelor's degree or a master's/PhD degree. However, among highly educated entrepreneurs, men are twice more likely to have graduated from an elite engineering school.

Female entrepreneurs have less industry experience and are less likely to have already founded a start-up: 65% of men indicate having at least three years of experience in the industry before starting up, while only 53% of women do. 33% of male entrepreneurs have already founded a start-up, whereas only 22% of female entrepreneurs have previous entrepreneurial experience as a founder.

Looking at the motivations to start, most of the entrepreneurs prefer to stay small. Women are even more likely to emphasize this point (75% versus 69% of men). Specifically, the average entrepreneur's main motivation is indeed to become independent and to a lesser extent, entrepreneurs choose this lifestyle because of a taste for entrepreneurship (45%). I find that women are less likely to indicate that they start because they have a taste for entrepreneurship, but are more likely to indicate that they founded their companies because of an opportunity. However, I find that there is no significant difference between the number of male and female

entrepreneurs who state that they found their companies because they want to explore a new idea (16%). Closely related to motivations, in my sample, 71% of entrepreneurs are married or in a relationship, and 56% have children. Regarding the composition of the founding team, I find that female and male entrepreneurs equally likely to start on their own (73%) or in a team with a relative (4.6%). Women are more likely to launch a new business with their spouse than men (12% versus 10%) and less likely to start in a team with business associates (11% versus 14%). Finally, consistent with Bordalo et al. (2017), women are less optimistic than men (21% versus 30%), at least according to the Landier and Thesmar (2008)'s measure.

[Insert table 1.2 here]

**What makes a female entrepreneur?** Table 1.A8 examines differences between male and female entrepreneurs in a multivariate framework, within the same sector at the same point in time. The average female entrepreneurs are significantly more educated but less experienced. Women are 5% more likely to have a bachelor's degree and 2% more likely to hold a Master's or PhD degree. However, they are 6% less likely to have a significant industry experience. In addition, female entrepreneurs are 8% less likely to have already founded another start-up in the past than male entrepreneurs. In sum, I find that female-start-up founders are less likely to start from a desire for independence or taste for entrepreneurship, but their entry is driven by perceived opportunities. In columns (3) and (4), I examine these differences in characteristics between male and female entrepreneurs who received succeeded at raising external equity. Even though the education and experience differences remain the same, differences regarding motivations of entrepreneurs whose the start-up is backed with external equity, are not statistically different from zero.

### **Are male- and female-founded start-ups different?**

Table 1.2 also reports the financing sources available to start-ups at creation. I find that 51% of entrepreneurs use external financing, and in particular bank loans (Robb and Robinson, 2014). 31% of entrepreneurs use bank loans, 13% use personal loans and 10% rely on non-bank loans. Regarding external equity, investors including VCs finance 3% of the start-ups in my sample. External equity financing is qualitatively

an important funding source for start-ups, as investors target start-ups with high-growth potential. In my sample, 1% of the male-founded start-ups and the 0.3% of female-founded start-ups are VC-backed. These figures correspond to approximately 1000 firms that receive VC funding among a sample of more than 92,000 start-ups.

My dataset also contains information on start-ups' employment size and performance from the year of creation to the five years onwards. At the end of the year of creation, 76% of firms in my sample do not have any employees. However, 2.5% of them have 6 or more employees. Significant differences exist between male- and female-founded start-ups. Although female entrepreneurs start smaller start-ups, the difference is smaller for the top bucket of the employment distribution, "11 or more employees". The average female-founded start-up is also smaller in terms of asset size and has a higher ratio of tangible assets. Considering unconditional means of performance, new firms founded by female entrepreneurs are less profitable over the first five years after creation than comparable start-ups founded by male entrepreneurs along three performance dimensions (sales, EBIT, and ROA). In addition, female-founded start-ups are incorporated in bigger and more feminized sectors, but their environments are on average not less competitive (measured by the Herfindahl index).

[Insert table 1.A8 here]

## 1.5. Main Results

### 1.5.1. Gender funding gap within sectors

To examine whether female entrepreneurs are at a disadvantage when raising financing, I compare the external equity fundraising success of female and male start-up founders within a sector at the same point in time. I control for various biographical characteristics that are arguably correlated with the probability of raising capital. Standard errors are clustered at the 4-digit SIC level to account for variation within. Figure 1.2 plots the findings and table 1.3 reports the results.

[Insert figure 1.2 here]

I find that female entrepreneurs are approximately 27% ( $=-0.0055/0.03$ ) less likely to use external equity relative to similar male-founded start-ups within the same sector, year, county (column 1). In column 2, I introduce dummy variables that account for education, prior industry and entrepreneurial experience. I find that entrepreneurs who hold a master's or a PhD are significantly more likely to use and have access to external equity financing relative to those who have not completed any higher education. The results also show that investors value prior entrepreneurial experience. While serial entrepreneurs are 1.3 percentage point more likely to raise external equity, prior industry experience does not convey any significant advantage in this regards. Controlling for these human capital variables do not fully explain the funding gap, which is of 19%. In columns 3 and 4, I focus on VC financing, and I find robust evidence that female-founded start-ups are 23-29% ( $=-0.0029/0.01$ ) less likely to receive VC.

Panel B presents a similar specification for alternative financing sources available at start-up creation. I find that gender does not explain the use of bank loans (column 1) and subsidies (column 4). However, the results show that female entrepreneurs are more likely to rely on personal loans and non-bank loans (columns 2 and 3). Besides, I find that the use of bank loans is positively related to citizenship, industry experience, and undergraduate education. However, more educated entrepreneurs do not seem to be more likely to rely on bank debt. Conversely, younger individuals and serial entrepreneurs are less likely to rely on bank debt, which seems to suggest that individuals who are likely to rely on external equity, are less likely to use alternative financing. In addition, the results show that asset tangibility is an important predictor of the use of external debt, whereas it is not significantly related to the use of external equity. This finding is consistent with the idea that banks focus on the quality of the collateral, as opposed to equity investors who learn from entrepreneurs' profiles.

Overall, controlling for observable human capital, biographical and start-up characteristics, I find that female entrepreneurs are less likely to contract with equity investors than similar male start-up founders within the same sector at the same point in time. This finding suggests that female entrepreneurs are *on average* at a disadvantage when it comes to raise external equity. I do not find a similar effect for

alternative financing sources. Male and female-founded start-ups are equally likely to use bank debt and subsidies. In addition, they are actually more likely to rely on personal debt and non-bank loans. These findings are consistent with the idea that capital structure and the use of different sources are equilibrium outcomes.

[Insert table 1.3 here]

### 1.5.2. The effect of gender stereotypes

If my previous results about external equity financing are driven by investors' preferences for male entrepreneurs, investors would systematically underfund female-led start-ups regardless of the sector. In contrast, if investors' beliefs about gender drive the results, then the marginal impact of female-dominated sectors on the likelihood of female-led start-ups accessing external equity finance should be positive. This section tests this hypothesis by interacting the entrepreneur's gender with measures of gender-dominated sectors.

The approach consists of identifying sectors in which female entrepreneurs represent the dominant group. I propose four measures of female-dominated sectors. First, to fit the prediction of the theoretical framework, I use the percentage of female entrepreneurs among the population of newly created firms by sector. A female-dominated sector is captured by a dummy variable that takes a value of one if more than 50% of those firms are founded by a woman.<sup>28</sup> The second measure relies on the percentage of female business owners for the population of newly created firms; the third, on the percentage of female business owners in the all population of firms; and the fourth, on the percentage of female CEOs. This set of regressions includes sector, cohort-year and county fixed effects, in addition to the human capital and start-up controls. I also control for time-varying sectoral characteristics, such as sector size, the Herfindahl index, and frequency of each gender by industry. This last variable ensures that the estimates are not picking up a mechanical relationship between the percentage of funded entrepreneurs by gender and their representativeness by industry. Figure 1.3 plots the unconditional means of fundraising success by gender between male- and female-dominated sectors.

---

<sup>28</sup>In unreported tables, I modify the threshold of 50% to 45% and 55%. Further, I exclude sectors including between 45% and 55%, and I exclude time-varying gender-dominated sectors. The results are broadly unchanged.

[Insert figure 1.3 here]

Table 1.4, panel A shows that although female entrepreneurs are significantly 23% ( $-0.0069/0.03$ ) less likely to raise external equity in male-dominated sectors, in female-dominated sectors, they are 3.6% ( $=0.0080-0.0069/0.03$ ) more likely to raise external equity relative to male entrepreneurs.<sup>29</sup> Relative to female entrepreneurs in male-dominated sectors, female entrepreneurs in female-dominated sectors are 26% ( $=0.008/0.03$ ) more likely to raise external equity. The effect is robust to the three above alternative definitions of female representativeness. The evidence suggests that investors are not systematically biased against women as taste-based discrimination would predict (proposition 1.2.3). In contrast, the asymmetric entrepreneurs' funding outcomes across sectors are consistent with context-dependent stereotypes, which predicts that investors are more likely to finance entrepreneurs when they belong to the majority group of an industry (proposition 1.2.5).

[Insert table 1.4 here]

In table 1.4, panel B, I examine entrepreneur fundraising success for alternative financing sources. As opposed to equity-financed start-ups, I do not find asymmetric entrepreneurs' funding outcomes by gender across sectors for other types of funding. Female entrepreneurs do not substitute their financing needs with alternative external financing sources, i.e., bank loans (column 2) and subsidies (column 4).

### 1.5.3. Future corporate outcomes

My approach of comparing fundraising success of female- and male-founded start-ups across male- and female-dominated sectors enables me to isolate belief-based discrimination from a pure taste-based discrimination story. However, this approach does not help to disentangle between the case where investors hold correct beliefs about gender (i.e., statistical discrimination) from the case where investors hold biased beliefs about gender (i.e., stereotypes). From the point of view of equity investors, it could be perfectly rational to finance male-founded start-ups more often in male-dominated sectors than female entrepreneurs in those sectors and than male

---

<sup>29</sup>Although the economic effect is positive, the effect is only statistically different from zero at 10%.

entrepreneurs in female-dominated sectors if male entrepreneurs' true average abilities are higher than those of female entrepreneurs. Similarly, it could be perfectly rational to finance female-founded start-ups more in female-dominated sectors than in male-dominated sectors if they are simply more able in those sectors and more able than men in female-dominated sectors.

Therefore, my approach consists of an outcome test that compares future corporate outcomes of successfully funded start-ups in male- and female-dominated sectors. If investors were rational and selected entrepreneurs according to the true average abilities of their gender group, we should not observe any systematic gender differences in future corporate performance within and across sectors. In contrast, the context-dependent stereotype view predicts that the marginal minority entrepreneur backed with equity performs better relative to entrepreneurs who belong to the majority group as the bar to be backed with equity was set higher for minority entrepreneurs than for majority entrepreneurs. The context-dependent stereotypes view also predicts that entrepreneurs should perform better in sectors in which they belong to the majority group as opposed to sectors in which they constitute the majority group for the same reason. Regarding future corporate performance of start-ups, I consider employment, sales over assets, the percentage of sales realized abroad, the returns over assets and survival measures from one year after creation to the five years onwards.

My results reported in table 1.5 panel A show that conditional upon being backed with equity, female-founded start-ups in male-dominated sectors perform marginally better than female-founded start-ups in female-dominated sectors along all measures of performance. However, the results are more contrasted when I compare the performance and development of male and female entrepreneurs within male-dominated sectors. On the one hand, female-founded start-ups employ more employees (column 1), use their assets more efficiently (column 2) and sell more abroad than male-founded start-ups in male dominated sectors (column 3). Specifically, the results show that female-founded start-ups in male-dominated sectors employ on average 0.7 more employees than male-founded start-ups in male-dominated sectors, up to five years after receiving external equity (column 1). Female-founded start-ups in male-dominated sectors outperform those in female-dominated sectors, as they

employ 3.7 more employees. Female-founded start-ups' use of assets, captured by the ratio of sales over assets, is also more efficient ( $0.7132-0.5700>0$ ) than their male counterparts within male-dominated sectors (column 2). On the other hand, female-founded start-ups have a lower returns over assets (ROA) relative to their male counterparts within male-dominated sectors ( $0.0505-0.0644<0$ ). However, the entrepreneurship literature argues that growth and development measures are better suited for start-ups the ROA that often have unstable net income in the first years of operation (Da Rin, Hellmann and Puri, 2013; Puri and Zarutskie, 2012).

The finding in column 3 indicates that female entrepreneurs in male sectors realize a larger part of their sales abroad relative to male entrepreneurs in these sectors ( $0.0271-0-0.0207$ ) and relative to female entrepreneurs in female-dominated sectors ( $0.0271$ ). Finally, results in columns 5 and 6 suggest no differences in the likelihood of surviving at least 3 and 5 years between female and male-founded start-ups within and across sectors.

In appendix table 1.A19, I look at male and female-founded start-ups that successfully exited either by IPO or M&A.<sup>30</sup> I find that start-ups founded by a female entrepreneur are more likely to go public in male-dominated sectors than male entrepreneurs, but also than female-entrepreneurs in female-dominated sectors (table 1.A19, column 1). I do not find any significant difference between male and female-founded start-ups in the likelihood to be acquired and go bankrupt at any point in the future, though (table 1.A19, columns 2-3).

Table 1.5 panel B examines the performance of male entrepreneurs in female-dominated sectors. The results indicate male entrepreneurs out-perform female entrepreneurs within female-dominated sectors along all measure of performance. Specially, male-founded start-ups that received external equity employ 1.5 more employees than similar female-founded start-ups in female-dominated sectors (column 1), they utilize their assets more efficiently (column 2), they sell more abroad (column 3) but they also have worse returns on assets (column 4) than their female counterparts in female-dominated sectors. Note that male entrepreneurs as a minor-

---

<sup>30</sup>Note that exits by IPO or M&A are relatively rare events (101 exists by IPO and 455 exists by M&A) and would not allow enough within sector variation within sector if I would include sector fixed-effects in the regression models. For this reason, models in appendix table 1.A19 do not include sector and cohort-year fixed effects and report the interaction of the start-up's gender with the gender-dominated sector of operation on the likelihood to exit by IPO (column 1) or by M&A or going bankrupt (columns 2-3).

ity also perform worse than those in male-dominated sectors (column 4). Finally, in appendix table 1.A19 panel B, I also find that male entrepreneurs in female-dominated sectors are more likely to exit by IPO relative to female entrepreneurs but also relative male entrepreneurs in other sectors. However, there is no significant differences in the likelihood to be acquired.

Overall, I show that entrepreneurs who belong to the minority group outperform entrepreneurs who belong to the majority group within sector, in terms of employment size, efficient use of assets, sales realized abroad and likelihood to go public. In addition, these minority entrepreneurs also perform better than entrepreneurs within the same gender across sectors where they constitute the majority group. Two notable exceptions to these findings are performance measured in terms of ROA and survival. First, entrepreneurs who belong to the majority group have a higher ROA. Second, I do not find that the minority group survive significantly longer than the majority group. An interpretation of this result could be that the ROA is a measure of performance that suit more mature firms as opposed to growth measures in terms of employment and sales (Puri and Zarutskie, 2012). Besides, survival measures reflect an equilibrium between the first and second moments of performance, such that a firm that fails does not necessarily only reflect bad management but also a risk taking attitude that is expected from start-ups that are backed with equity, to grow.

[Insert table 1.5 here]

**Risk-neutrality hypothesis?** However, the better performance of the minority group over the majority group within and across sector could also be rationalized by the fact that minority entrepreneurs start riskier businesses. Hence, one needs to assume risk-neutrality to interpret the performance results in table 1.5. I test the validity of this assumption in columns 7 and 8 of table 1.5. I consider two proxies for risk: the volatility of firm's employment size and the firm's volatility of sales from the year of creation to five years onwards.<sup>31</sup> The results show that minority-led start-ups are not significantly riskier than start-ups founded by entrepreneurs who belong to the majority group both within and across sectors, for the two measures

---

<sup>31</sup>Note that firms that it is not possible to compute on firms that do not survive more than one year, which explains the drop in the number of observation from 2705 to 2334.

of risk.

**Marginal or average entrepreneur?** To establish these facts, I rely on OLS estimates. A concern could be that standard OLS estimates recover the performance of the average entrepreneur from each group, and not of the marginal entrepreneur. As a result, comparisons based on OLS estimates will not recover the true level of bias in external financing decisions unless one is willing to assume that there is an identical distribution of performance across groups (Dobbie et al., 2018). I test for this assumption by visual inspection of the OLS residuals distributions by gender and sector. Figures 1.4 plot the findings for the employment regression estimates and the sales over assets regression estimates. The Residuals' distributions of male and female entrepreneurs' performance in male-dominated sectors are largely comparable both in terms of mean and standard deviations. However, in female-dominated sectors, the mean of female entrepreneurs' residuals' distributions seem higher and the variance lower than those of male entrepreneurs.

[Insert figures 1.4 here]

**Alternative measures of performance.** The literature that usually establishes the performance of PE and VC backed firms usually rely on exit outcomes, such as IPOs, M&As and bankruptcy. This choice is often motivated by the availability of this information in commercial database (Da Rin, Hellmann and Puri, 2013). In the sample, 6% of the firms that received PE or VC investment go public, 23% get acquired, 8% go bankrupt and the rest stays private, at least til April 2019, when the data were collected. In appendix table 1.A19, I replicate the performance findings using these measures of performance.<sup>32</sup> I find that female entrepreneurs in male-dominated sectors are significantly more likely to exit by IPO than male entrepreneurs in those sectors. Similarly, male-founded start-ups are significantly more likely to go public than female-founded start-ups within female-dominated sectors (columns 1). In contrast, I do not find that male and female-founded start-ups have a different probability to be acquired or to go bankrupt (columns 2 and

---

<sup>32</sup>Note that the specifications do not include fixed effects as IPOs are relatively rare events. With fixed effects, the coefficient would be identified on the very few industries that include both a male and female-founded start-ups that exit by IPO within the same industry, the same year.

3). To sum, the better performance of the minority group seems to be driven by a few firms that will go public.

#### **1.5.4. Mechanism: Higher bar for minorities or selected pool of entrepreneurs?**

In the previous section, I argue that the better performance of entrepreneurs who belong to the minority group is consistent with the idea that the bar is set higher for minorities. An alternative interpretation is related to entrepreneurs' self-selection into sectors. Female entrepreneurs who self-select into female-dominated sectors would be of a lower-ability type than those who self-select into male-dominated sectors. This could be, for instance, the case if entrepreneurs enjoy private benefits of starting in a sector where they fit the representative gender. Table 1.6 panel A supports this intuition. Looking at the population of entrepreneurs, I find that female entrepreneurs who self-select in female-dominated sectors are on average of worse quality than those who self-select in male-dominated sectors along several observable dimensions in the data, such as education, experience and motivations.

In table 1.6 panel B focuses on the subsample of entrepreneurs who successfully raised external equity. I do not find any significant differences anymore between male and female entrepreneurs who self-select into male- or alternatively female-dominated sectors regarding education in elite engineering schools and prior industry or entrepreneurial experience. Regarding *ex ante* motivations for entry into entrepreneurship (new idea, taste, opportunity, independence), successfully funded female entrepreneurs who choose a female-dominated sector do not seem to be significantly different from those who choose a male-dominated sector. However, those female entrepreneurs who prefer a male-dominated sector over a female-dominated sector are more likely to state in the survey that they have the ambition to grow and they are more likely to have optimistic beliefs regarding their future growth (Landier and Thesmar, 2008). Overall, the evidence suggests that women who passed the bar of being selected by investors in a male-dominated sector are not *observably* different in terms of experience and motivation for entry than those who opted to start in a female-dominated sector.

[Insert table 1.6 here]

## 1.6. Robustness Tests

In this section, I test whether human capital helps minority entrepreneurs to be less subject to discrimination from external equity investors. Second, I test whether my previous results correlate with cross-sectional variations in the composition of founding teams. Third, I test the robustness of my previous results to various factors arguably correlated with entrepreneurs' abilities, such as founders' motivations and optimism about the future start-up's growth. Fourth, I show that investors are subject not only to gender stereotypes but also to age stereotypes.

### 1.6.1. Human capital

Are more educated and more experienced entrepreneurs more likely to be selected by investors? And in particular, does this help minority entrepreneurs to access finance? In this section, I interact the human capital variables with the gender and the variable female-dominated. All specifications include county, sector, and cohort-year fixed effects, as well as human capital control variables and start-up controls. Standard errors are clustered at the 4-digit French SIC level. Table 1.7 reports the results. Consistent with the main findings, female entrepreneurs are still on average less likely to raise external equity than male entrepreneurs. However, they balance their disadvantage in female-dominated sectors in which they are actually more likely to raise external equity than male entrepreneurs.

The results also show that higher education and education from elite schools are positively associated with the likelihood of raising external equity. In column 1, male entrepreneurs who hold a master's degree have 0.0106 percentage points more to be backed with equity relative to entrepreneurs with no degree or only undergraduate education. Female entrepreneurs who graduated from a master's or PhD degree are also more likely to access external equity, although the effect is lower ( $0.0106 - 0.0063 = 0.4\%$ ). Besides, entrepreneurs who graduated from an elite engineering school are also more likely to use external equity (column 2). However, is the effect of graduating from an elite engineering school equally favors male and female entrepreneurs. Note that 6% of male start-up founders and 3% graduated from one of these school. Finally, the results show that both male and female

engineers entrepreneurs are also more likely to have received external equity in female dominated sectors.

Regarding prior industry and entrepreneurial experience, I still find that industry experience is not significantly related to the likelihood of accessing external equity, this finding being consistent across genders and sectors. Besides, I find that being a serial entrepreneurs increases the likelihood of getting access to external equity financing by 0.0161 percentage points for male entrepreneurs and by 0.0029 percentage points for female entrepreneurs. Overall, the results suggest that human capital measured by higher education and entrepreneurial experience significantly increases the chances of accessing equity financing for both male and female entrepreneurs, the effect being more moderate for the later.

### **1.6.2. Team and family**

Teams of entrepreneurs represent 29% of newly founded firms in France. In this section, I test the effect of starting as a stand-alone entrepreneur (column 1) as opposed to within a team (column 2). According to the survey's rules, the person who completes the questionnaire is supposed to be the business owner. In table 1.8, I find that team-founded start-ups led by a female entrepreneur are significantly (-48%=-0.0144/0.03) less likely to raise external equity relative to team-founded start-ups led by a man. Women as team leaders are indeed less likely to raise external equity than the ones who start on their own. Stand-alone female entrepreneurs are -16% (= -0.0049/0.03) less likely to raise external equity. Regarding gender stereotypes, female-led teams and stand-alone female entrepreneurs both benefit from starting in female-dominated sectors as they seem to be able to partly or fully balance the disadvantage of being a women.

Next, I examine the composition of entrepreneurial teams. The founding team can be formed by spouses (column 3), family members (column 4), or associates and business partners (column 5). I find that teams formed by spouses do not experience discrimination due to gender. The evidence suggests that gender matters in the eyes of investors and that the initial disadvantage of being a female entrepreneur is balanced when a women starts with her spouse. Besides, I find that female-led start-ups formed by family members associates are less likely to be funded in male-

dominated sectors and do not benefit from starting in a female-dominated sector. Thus, female-led start-ups that include professional associates have fewer chances to contract with external equity investors compared to male-led professional start-ups.

Finally, I consider the effects of entrepreneurs' marital status and I find that single female start-up founders are about three times less likely to raise external equity. However, they benefit from starting in a female-dominated sector. Overall, table 1.8 shows that investors perceive female-led teams less favorably than comparable male-led teams. Nevertheless, teams formed by spouses do not experience such negative effects. The evidence suggests that investors value the presence of men in entrepreneurial teams.

[Insert table 1.8 here]

### 1.6.3. Motivations and optimism

Female entrepreneurs may start a company for different reasons than their male counterparts. While a large proportion of entrepreneurs indicate that they create a start-up to enjoy the private benefits of being their own bosses, external equity investors are looking for ambitious and high-growth oriented entrepreneurs. In this section, I study the effects of initial motivations and ambitions about future corporate development, both stated *ex ante* by the entrepreneur, on the use of external equity finance.

**Ambition to grow.** Table 1.9 column 1 presents the effect of entrepreneurs who have a strong ambition to grow relative to entrepreneurs who simply want to create a job for themselves on the probability of raising external equity. Start-up founders who have the ambition to develop their firms are 40% ( $=0.0119/0.03$ ) more likely to raise capital relative to those who start to stay small. However, those female entrepreneurs are 0.0091 percentage points less likely to raise external equity relative to male entrepreneurs who state the same ambition.

**Initial motivations.** Columns 2 to 5 examine whether female start-up founders remain at a disadvantage even after controlling for their main initial motivations. I find that a new idea, a taste for entrepreneurship, or the perception of an opportunity are significantly and positively related to the likelihood of raising external equity. No difference exists by gender across sectors, except for entrepreneurs in female-

dominated sectors who declare that they have a taste for entrepreneurship. External equity investors negatively perceive independence as a motivation to become an entrepreneur but this is more favorably appreciated for female entrepreneurs, with no significant difference across sectors.

[Insert table 1.9 here]

**Optimistic entrepreneurs.** Given the risk profile of VC-backed projects (they have high levels of risk and right-skewed distributions), optimistic entrepreneurs may be more willing to seek VC financing than realistic entrepreneurs. Therefore, I wonder whether a relation exists between fundraising success and gender when controlling for optimism. I find that entrepreneurs who hold optimistic beliefs are 28% more likely to use external equity financing, but there is no significant difference across sectors. Overall, the results suggest that highly motivated entrepreneurs are more likely to raise external equity for their start-ups both in male and female-dominated sectors. However, highly motivated female-entrepreneurs seem to benefit less than men from these high growth orientation, taste for entrepreneurship and optimism, with the exception of female entrepreneurs who started because of a new idea or an opportunity.

#### 1.6.4. The effect of age stereotypes

Investors may hold stereotypes not only about gender but also about other entrepreneurs' traits such as age for instance.<sup>33</sup> In this section, I test the marginal effect of young sectors on the likelihood that old entrepreneurs (who are 50 years or older) use external equity financing (including VC) and alternative financing sources. Young sectors are flagged as sectors in which the median CEO's age is younger than 40 years.

Table 1.10 reports the results. First, I find that being an old entrepreneur does not impact the probability of using external equity finance (column 1). However, it negatively affects the likelihood of raising bank debt, getting a personal loan or obtaining a grant or a subsidy (columns 2, 3 and 4 respectively). Second, I find

---

<sup>33</sup>I also tested the effect of nationality stereotypes, i.e., being French or not and being European or not, but I did not find similar results. One potential explanation is that there is not enough variation in the data, foreigners never constitute the dominant group of an activity both in terms of numbers of entrepreneurs, CEOs or employees.

that old entrepreneurs are significantly worse off than younger entrepreneurs when it comes to accessing external equity in sectors in which the dominant age group is being younger than 40 years old. This finding is consistent with investors behaving according to stereotypes. External equity investors associate a negative value to entrepreneurs who differ from the dominant type in the environment. Nevertheless, old entrepreneurs in young sectors are not significantly less likely to access bank debt and equity grants than young entrepreneurs.

[Insert table 1.10 here]

## 1.7. Further Alternative Explanations

A concern in survey-based datasets is that entrepreneurs may make mistakes when they self-report their financing sources available at creation. I replicate my main results out of sample using extracts of VC and other PE deals from the VentureXpert database linked with the French administrative data. In addition, I take advantage of having the investor's identity to test whether homophily could explain why start-ups run by an entrepreneur who belongs to the minority gender group outperform. I retrieve the gender of PE and VC fund managers by linking those investment firms with the French matched employer-employee dataset, and I test whether female entrepreneurs' performance can be explained by being advised by a female investment partner at a PE and VC firm.

### 1.7.1. Out of sample replication

#### Construction of the dataset

The dataset in this section consists of the commercial database Thomson VentureXpert merged with French administrative data (tax files and employer payrolls). VentureXpert is a proprietary database that contains information on PE deals, including VC, unspecified PE, leveraged buyout (LBO) deals, and financing by angel investors. I discard LBO deals.<sup>34</sup> The matching between commercial databases and

---

<sup>34</sup>I also exclude companies with missing names or addresses because they cannot be matched to French administrative data. Note that my extracts from VentureXpert include only angel- and PE-VC-financed companies located in France.

the universe of French administrative data relies on a Python web-crawler (see appendix 1.A21 for more details). I identify firms in the tax files as PE-financed if they can be matched to extracts of VentureXpert, which identifies firms that receive their first round of PE financing between January 2002 and December 2016. Thus, I can observe financial performance and employment composition from firm creation to exit. The final sample contains more than 9,000,000 firm-year observations.

For this study, it is important that I can identify the company CEO's gender. The French employer payrolls database provides a 4-digit occupation code for each employee (414 different occupations), including CEOs for different firm sizes starting at 10 employees. The highest paid employee is considered to be the CEO for firms with fewer than 10 employees. I code a dummy variable *Female* if a female CEO runs the company. In addition, I compute firm-level percentages of female employees, female engineers, and female executives. A similar procedure applied to PE and VC investment firms allows me to identify the gender of their general partners (GP) and potential board members, as the top five most compensated employees.<sup>35</sup>

### **Gender stereotypes and types of investment firms**

Appendix table 1.A18 replicates the results of table 1.4 for all PE, VC, corporate venture capital (CVC), and angel investment deals in France from 2002 to 2015. The approach consists of comparing deals within the same 4-digit French SIC sector, the same size bucket and the same cohort-year.<sup>36</sup> In addition, specifications include the firm's tangible ratio and the logarithm of total assets. My results show that female-led companies are significantly less likely to contract with PE and VC investors. My results show similar trends for CVC funds. I do not find a similar effect for angel investors, as this type of equity investor may have access to supplementary private information about the firm.

To test the effects of stereotypes, I interact the CEO's gender with a dummy variable identifying female-dominated sectors. My results show that female entrepreneurs in female-dominated sectors are significantly more likely to receive PE

---

<sup>35</sup>Note that the identity of the investment firm is often not available for angel investors in VentureXpert.

<sup>36</sup>Cohort-years are defined as the -1, +1 years around a firm's creation date. Size buckets consist of 10 categories based on the number of employees: zero, [1;4], [5;9], [10;19], [20;49], [50;99], [100;249], [250;499], [500;999], and more than 1,000.

and VC. In contrast, angel investors do not seem to display stereotypical beliefs-based behaviors. One reason could be related to the fact that angel investors have superior private information about the entrepreneurs, since relationships are informal.

### **1.7.2. Investor gender, homophily and future corporate performance**

I test the effect of the external equity investors' gender on future corporate performance. The test is conducted on a sample of PE and VC deals only as the angel investors' identity and investors' gender cannot be identified in most cases. The approach consists of comparing successfully funded firms 5 years before and 5 years after being granted external equity and comparing the effect of having a general partner at the VC firm that is female. I consider five measures: the number of employees, the logarithm of the number of employees, the percentage of female employees, the ratio of sales over assets and the ROA. All models include 4-digit SIC sectors, county and year fixed effect.

Table 1.12 suggests that firms grow more in terms of employment after receiving a first round of PE or VC investment (columns 1 and 2). However, their performance, measured by the ratio of sales over assets and ROA, seem to deteriorate in the post period (columns 4 and 5). In addition, performance and growth of female-led firms in the post period are not significantly different from those of male-led firms. Moreover, my results show that female investors are positively related to better corporate performance. Equity backed start-ups selected by a female partner have more employees and use their assets more efficiently (columns 1, 2 and 4). However, the coefficient of interaction between the investor's gender and the post investment period is not significantly different from zero, which suggests that female investors do not treat invested firms better than male investors, but seem better at selecting the best firms in the first place, though. Finally, the coefficient of the triple interaction between the entrepreneur's gender, the investors' gender and the post investment period is not significantly different from zero, suggesting that female investors are better at advising female entrepreneurs than male investors.

[Insert table 1.12 here]

The evidence in table 1.12 suggests that future corporate outcomes are not influenced by homophily between an entrepreneur and her advisor. However, the fact that a few investment firms are managed by female fund managers (8.9% of my sample, see appendix table 1.A17) may not allow enough variation on the supply side to identify a significant relationship. Homophily, also called in-group bias, requires sufficient groups from both sides to have a significant impact Jannati et al. (2016). To address this concern, I replace test the robustness of my previous findings to a variable that captures to some extends the feminization of the board at the VC firm. I define a dummy variable that takes the value one if at least one of the five most compensated employees at the VC is a women. 65% of VC firms in my sample have a women among their top employees. In appendix table ?? reports the results. I still find that investors' gender has very little influence on performance of selected start-ups (columns 4 and 5), with a notable exception for the log of the firm's employment size that is higher in the case a female entrepreneurs advised by a VC board with at least one women in the post investment period (column 2). However, this finding is not robust when I consider the number of employees instead of the log number. Overall, homophily on gender between the entrepreneurs and her advisers do not seem to explain the better performance and corporate growth in the post investment period.

## 1.8. Conclusion

This paper examines the effect of gender stereotypes on the likelihood of raising PE and VC financing. My empirical analysis relies on the merging of different administrative datasets available at the French Statistical Institute. The final dataset contains detailed information about start-ups' characteristics and entrepreneurs' personal traits, and is not subject to the usual selection biases encountered in the entrepreneurship literature. One fact that emerges from the analysis is that female-founded start-ups are less likely to contract with equity investors within and across sectors. However, the novelty of this paper is to show that female entrepreneurs are not always at a disadvantage when it comes to contracting with financiers. In female-dominated industries, female-founded start-ups are equally to more likely to

raise capital relative to male-founded start-ups in female-dominated sectors, and are significantly more likely to raise capital than female-founded start-ups in male-dominated sectors. The evidence is consistent with investors who have context-dependent stereotypes as opposed to rational beliefs and preference toward male entrepreneurs. I isolate the effects of stereotypes from competing explanations by developing a theoretical framework that generates distinct empirical predictions for the cases where investors are systematically biased against a gender, are rational in their discrimination, or have context-dependent stereotypes.

If taste-based discrimination were the primary driver of investors' behaviors, investors would systematically discriminate against the same gender group regardless of their abilities and sectors of incorporation. My results show that female entrepreneurs are not systematically at a disadvantage when raising capital. Differences in start-ups' initial size, asset tangibility, and performance, as well as gender differences in human capital, ex ante motivations, and optimism do not fully explain the observed gender gap. Furthermore, I find that start-ups founded by mixed teams that comprise spouses no longer experience the negative effect of being female-founded. The evidence suggests that equity investors pay attention to start-up founders' gender.

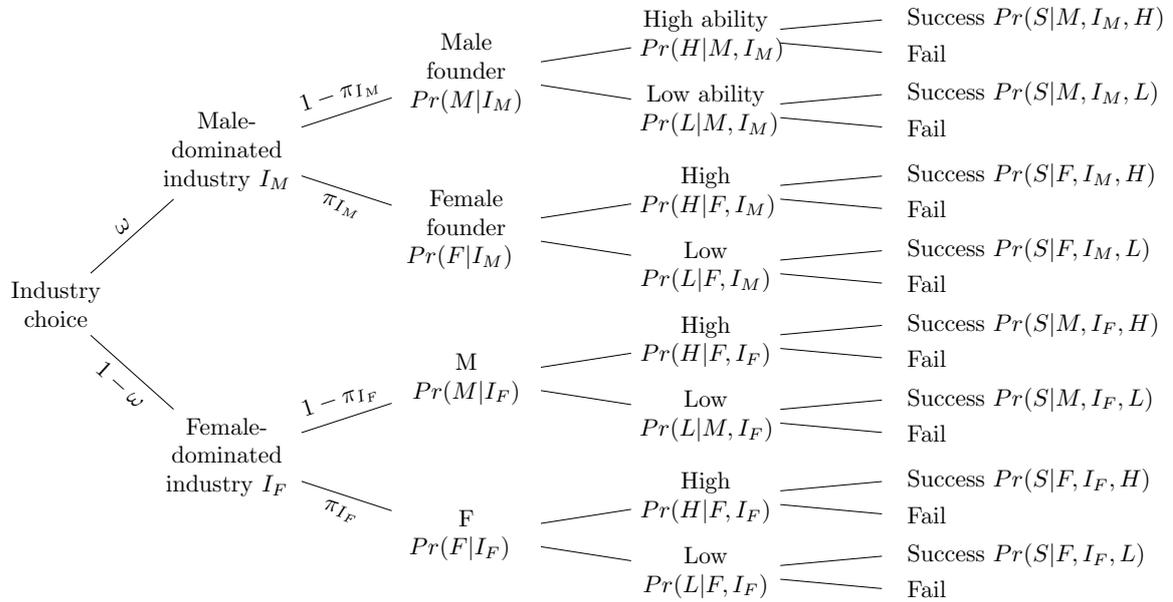
Furthermore, I find that conditional on receiving PE and VC financing, female-founded start-ups in male-dominated sectors perform better relative, not only to similar male-founded start-ups in those sectors, but also to female-founded start-ups in female-dominated sectors. The inverse is also true for successfully funded male-led start-ups in female-dominated sectors. Under the assumption that abilities are constant over time, this finding suggests that the ability bar for minority-led start-ups to be funded is set higher than for non-minority-led start-ups. As a result, a non-negligible share of low-ability entrepreneurs who belong to the dominant gender group is funded at the expense of minority entrepreneurs. This finding is consistent with the predictions of the stereotypical view as opposed to statistical discrimination rooted in rational expectations about genders.

Investors react differently to entrepreneurs' gender depending on whether they belong to the dominant gender group of a context. My study sheds new light on the negative relationship between female entrepreneurs, and the probability that

they successfully raise capital. I show that part of the gender funding gap arises from composition effects in the structure of the economy, and not only from exogenous and constant investors' preferences towards a specific gender. As female-dominated industries represent a lower share of the economy, male stereotypes dominate those that favor female entrepreneurs. Stereotypes are attenuated when genders are equally distributed among sectors. This study helps to rationalize policy interventions that aim to reach gender equality by increasing female participation in sectors in which they are underrepresented.

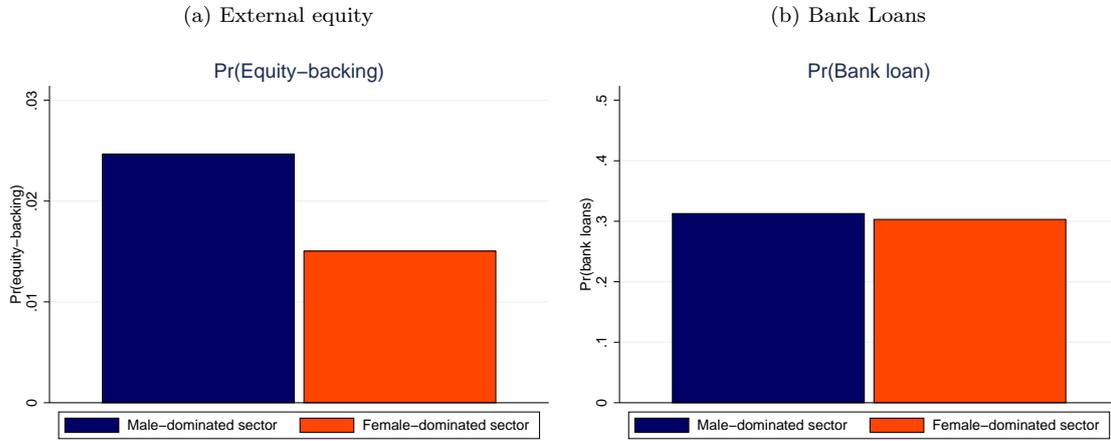
## 1.9. Figures and Tables

**Figure 1.1: Decision Tree of a Financier Who Faces a Population of Entrepreneurs**



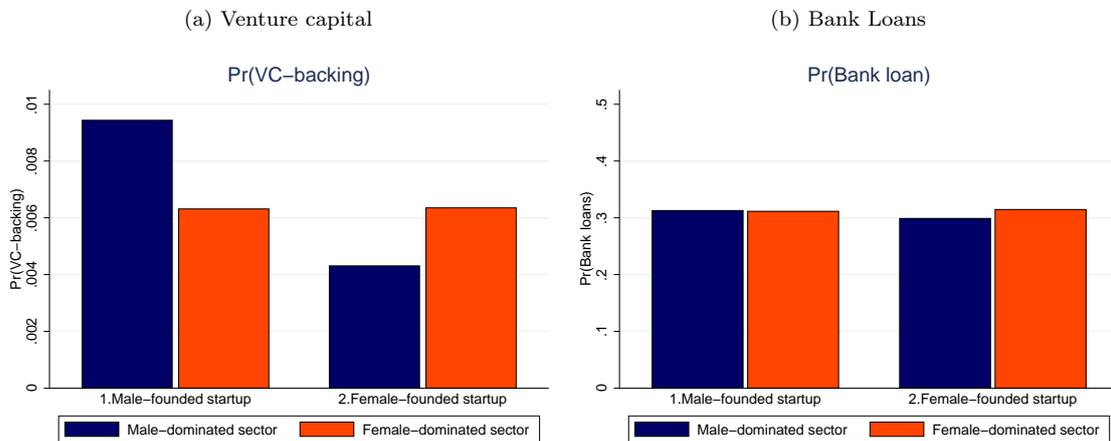
**Figure 1.2. Gender Funding Gap**

*Source:* SINE survey. This figures plot the unconditional means by founder gender group of start-up that receive VC or other external equity financing or bank loans.



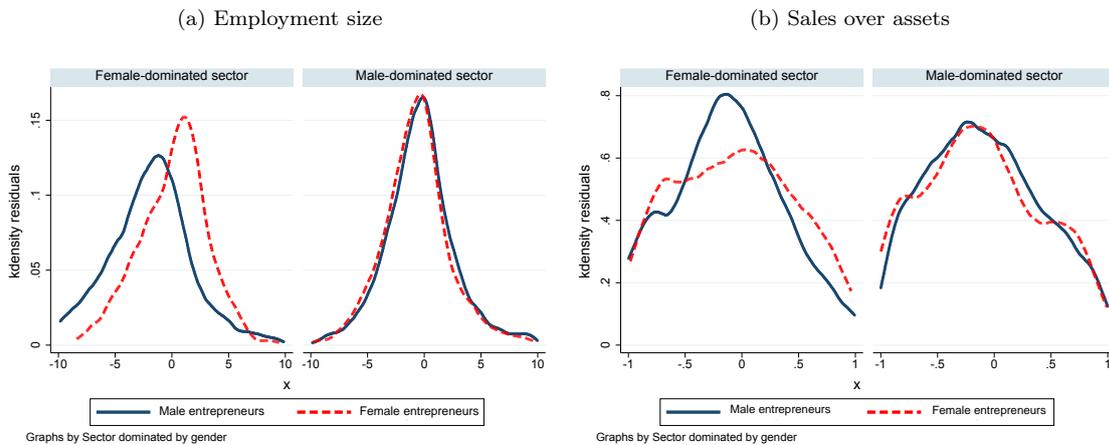
**Figure 1.3. Gender Funding Gap and Context-dependent Stereotypes**

*Source:* SINE survey. The figures plot the unconditional means by founder gender group and gender-dominated sectors of start-ups that receive VC financing or bank loans. A female-dominated sector includes at least 50% of new female-founded start-ups within a 4-digit French SIC sector.



**Figure 1.4. Residuals' Distributions of the Outcome Tests**

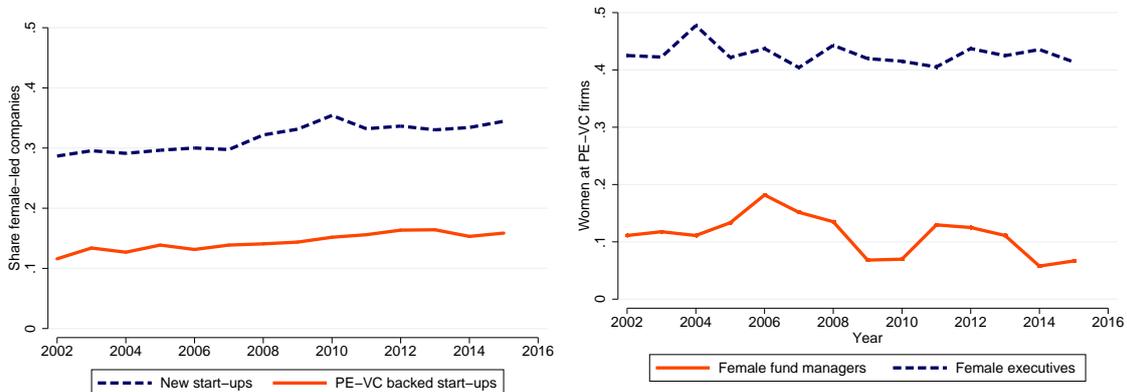
*Source:* SINE surveys and tax files. The figures plot residuals' distributions from the outcome tests (table tab-perf-future) and check their similarities between male and female start-up founders by sectors. The outcome tests are based on OLS estimates that could be biased unless the distributions of the outcome variable by gender group are similar. An outcome test consists in regressing future corporate employment size (panel A) and future reported sales (panel B) on the start-up founder's gender and a dummy variable male-dominated sector (equation 1.13). A male-dominated sector is the exact inverse of a female-dominated sector and includes at least 50% of new male-founded start-ups within a 4-digit French SIC sector.



**Figure 1.5. Female Representation in the PE-VC Industry**

*Source:* VentureXpert and employer payrolls. Figure (a) represents the share of female-led companies that receive PE or VC financing from 2002 to 2015. Figure (b) displays the share of PE and VC investment firms led by a female CEO (general partner), the share of female executives at PE and VC investment firms and the share of female workers at PE and VC investment firms

(a) Percentage of female-led PE- and VC-backed companies (b) Female representation at PE and VC investment firms



**Table 1.1. Male- and Female-founded Start-ups by Cohort**

*Source:* SINE survey. This table reports numbers of observations and percentages by groups of male and female-founded start-ups in 2002, 2006, 2010, and 2014.

	Cohorts				
	2002	2006	2010	2014	Total
Male	9,718	11,809	24,613	21,899	68,039
%	77	75	73	72	74
Female	2,896	3,911	9,067	8,533	24,407
%	23	25	27	28	26
Total	12,614	15,720	33,680	30,432	92,446

**Table 1.2. Entrepreneurs' Personal Traits and Start-ups' Characteristics**

*Source:* SINE surveys and tax files. *Sample:* Repeated cross-section of new firms founded in 2002, 2006, 2010, and 2014. The table presents entrepreneurs' biographical characteristics (panel A), motivation and optimism items (panel B), family structure and composition of the founding team (panel C), start-up's financing sources at creation (panel D), employment size at creation (panel E), balance sheet information from the creation year to five years onward (panel F), and sectoral characteristics (panel G). The mean and number of observations by gender group are reported as well as t-statistics and p-values of the mean differences between male and female entrepreneurs. Variable definitions and data sources are provided in Appendix C.

Variables	All	Male		Female		difference	t-stat
	Mean	N	Mean	N	Mean		
Panel A. Biographical characteristics							
Age $\geq$ 40	0.48	68,039	0.49	24,407	0.45	0.04***	(11.28)
French	0.91	68,039	0.90	24,407	0.93	-0.03***	(-13.55)
No degree	0.20	68,039	0.21	24,407	0.16	0.05***	(18.45)
High school	0.44	68,039	0.45	24,407	0.42	0.04***	(9.62)
Bachelor's	0.14	68,039	0.13	24,407	0.17	-0.04***	(-15.79)
Master's/PhD	0.22	68,039	0.21	24,407	0.26	-0.04***	(-13.76)
Elite engineering school	0.05	68,039	0.06	24,407	0.03	0.03***	(19.48)
Expert	0.62	68,039	0.66	24,407	0.53	0.13***	(35.16)
Serial	0.30	68,039	0.32	24,407	0.22	0.10***	(31.80)
Panel B. Initial motivations and optimism							
High-growth oriented	0.29	58,768	0.31	21,694	0.25	0.06***	(16.40)
<u>Motivation for entry:</u>							
Independence	0.61	68,037	0.62	24,407	0.60	0.02***	(6.18)
Taste	0.45	68,035	0.46	24,406	0.42	0.04***	(10.08)
Add earnings	0.24	58,318	0.25	21,511	0.22	0.03***	(10.13)
Opportunity	0.20	68,033	0.20	24,406	0.22	-0.02***	(-6.63)
New idea	0.16	68,033	0.16	24,406	0.17	-0.01***	(-4.09)
<u>Optimistic entrepreneurs:</u>							
Optimism at start	0.30	40,140	0.33	15,264	0.22	0.11***	(26.67)
Optimism at t+3	0.16	34,025	0.17	11,904	0.13	0.04***	(11.41)
Optimism at t+5	0.13	19,076	0.14	6,617	0.11	0.03***	(7.27)
Panel C. Family structure and team composition							
<u>Family structure:</u>							
Married	0.71	58,321	0.73	21,511	0.68	0.05***	(13.49)
Children	0.56	33,708	0.56	12,444	0.58	-0.02***	(-4.52)
<u>Team composition:</u>							
Alone	0.73	66,515	0.73	23,551	0.73	-0.00	(-0.90)
Spouse	0.11	66,513	0.10	23,550	0.13	-0.03***	(-10.65)
Relatives	0.05	66,512	0.05	23,550	0.05	-0.00	(-0.96)
Associates	0.13	66,514	0.13	23,550	0.11	0.03***	(11.41)

**Entrepreneurs' Personal Traits and Start-ups' Characteristics**  
(Continued)

Variables	All	Male		Female		difference	t-stat
	Mean	N	Mean	N	Mean		
Panel D. Financing sources							
External financing	0.51	68,039	0.51	24,407	0.52	-0.01**	(-2.07)
External equity:	0.03	68,039	0.03	24,407	0.02	0.01***	(10.10)
Venture capital	0.01	56,230	0.01	20,496	0.007	0.00***	(7.45)
Business equity	0.03	68,039	0.03	24,407	0.02	0.01***	(7.98)
Bank loans	0.31	68,039	0.31	24,407	0.30	0.01***	(2.81)
Personal loans	0.13	68,039	0.12	24,407	0.14	-0.02***	(-6.73)
Non-bank loans	0.10	68,039	0.09	24,407	0.11	-0.02***	(-9.47)
Subsidies	0.17	68,039	0.17	24,407	0.18	-0.01*	(-1.83)
Panel E. Employment size							
Employment size at start	0.69	67,039	0.73	23,907	0.59	0.15***	(5.96)
Zero	0.77	67,039	0.76	23,907	0.79	-0.03***	(-9.10)
1	0.11	67,039	0.11	23,907	0.10	0.01***	(2.63)
2	0.05	67,039	0.05	23,907	0.05	0.01***	(3.17)
3	0.02	67,039	0.03	23,907	0.02	0.00***	(3.78)
3	0.02	67,039	0.03	23,907	0.02	0.00***	(3.78)
4-5	0.02	67,039	0.03	23,907	0.02	0.01***	(6.14)
6-10	0.02	67,039	0.02	23,907	0.01	0.00***	(5.24)
11+	0.01	67,039	0.01	23,907	0.01	0.00**	(2.55)
Employment size at t+3	1.14	37,627	1.20	12,344	0.96	0.24***	(5.00)
Employment size at t+5	1.18	34,309	1.24	11,458	1.00	0.24***	(4.25)
Panel F. Balance sheet information and income statement (Panel 5 years)							
N. employees	1.61	257243	1.672	82508	1.406	0.27***	(9.31)
Sales over assets	17.25	298776	17.319	102428	17.061	0.26	(0.08)
% foreign sales	0.03	288714	0.031	99466	0.033	-0.00***	(-3.37)
roa	0.26	298776	0.251	102428	0.301	-0.05***	(-7.10)
Survival 3 years	0.86	298776	0.867	102428	0.841	0.03***	(19.75)
Survival 5 years	0.75	298776	0.761	102428	0.732	0.03***	(18.31)
Tangible/total assets	0.23	298776	0.224	102428	0.246	-0.02***	(-2.73)
Ln(total assets)	3.83	298776	3.884	102428	3.657	0.23***	(41.53)
Panel G. Sectoral characteristics							
Sector size (total sales M. euros)	18.56	1402	19.235	239	14.583	4.652**	(2.45)
N. start-ups	140	1414	148	242	96	52.59***	(2.64)
N. backed firms within sector	3.81	1414	4.20	242	1.54	2.67***	(6.34)
N. VC-backed firms within sector	0.70	1414	0.77	242	0.29	0.48***	(4.10)
Start-ups' characteristics within sector:							
Mean Sales (k euros)	8.99	1402	9.52	239	5.90	3.62**	(2.52)
Vol Sales (k euros)	61.83	1400	66.11	239	36.74	2.98***	(3.27)
Mean Employment size	28.48	1402	29.64	239	21.70	7.94***	(2.63)

**Table 1.3. Female Entrepreneurs, External Equity and Other Financing Sources**

*Source:* SINE surveys and tax files. *Sample:* Repeated cross-section of new firms founded in 2002, 2006, 2010, and 2014. The table reports linear probability model estimates and analyzes the effect of gender on the use of different financing sources. The dependent variables are as follows: panel A, columns 1-2, a dummy variable that is equal to 1 if the start-up receives VC or other external equity financing; panel A, columns 3-4, a dummy variable that is equal to 1 if the start-up receives VC financing; panel B, column 1, a dummy variable that is equal to 1 if the start-up uses any sources of external financing; panel B, column 2, a dummy variable that is equal to 1 if the start-up uses bank loans granted to the company; column 3, a dummy variable that is equal to 1 if the start-up is funded by bank loans granted to the entrepreneur; and panel B, column 4, a dummy variable equal to 1 if the start-up receives an equity grant. *Female* is a dummy variable that is equal to 1 if the start-up is run by a woman. *French* is a dummy variable that is equal to 1 if the entrepreneur is a French citizen. *Bachelor's* is a dummy variable that is equal to 1 if the entrepreneur has a bachelor's degree. *Master's/PhD* is a dummy variable that is equal to 1 if the entrepreneur has at least a five-year university degree. *Expert* is a dummy variable that is equal to 1 if the entrepreneur has at least three years of work experience within the sector. *Serial* is a dummy variable that is equal to 1 if the entrepreneur has previously founded a company. Start-up controls are included and comprise the incorporation status dummy variable, the ratio of tangible assets and the logarithm of firm's total assets. All models include county and 4-digit French SIC sector  $\times$  cohort-year fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

## Panel A. External Equity and Venture Capital

	External equity		Venture capital	
	(1)	(2)	(3)	(4)
Female	-0.0081*** (-6.76)	-0.0055*** (-4.66)	-0.0029*** (-5.13)	-0.0023*** (-3.89)
Age $\geq$ 40		0.0026** (2.58)		0.0010* (1.79)
French		-0.0018 (-0.94)		0.0008 (0.80)
Bachelor's		-0.0009 (-0.60)		-0.0020*** (-2.99)
Master's/PhD		0.0085*** (5.73)		0.0029*** (3.44)
Expert		0.0002 (0.20)		0.0002 (0.22)
Serial		0.0130*** (10.46)		0.0019** (2.42)
Incorporated		0.0156*** (11.68)		0.0045*** (5.82)
Tangible/total assets		-0.0018 (-0.70)		-0.0005 (-0.29)
Ln(total assets)		0.0038*** (8.21)		0.0010*** (3.59)
Sector $\times$ cohort FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
$R^2$	0.038	0.044	0.039	0.040
N	92338	92338	76725	76725

**Female Entrepreneurs, External Equity and Other Financing  
Sources**  
*(Continued)*

Panel B. Other Financing Sources

	External financing (1)	Bank loans (2)	Personal loans (3)	Subsidies (4)
Female	0.0063 (1.01)	-0.0029 (-0.63)	0.0069** (2.41)	0.0029 (0.71)
Age $\geq$ 40	-0.0364*** (-8.99)	-0.0406*** (-11.16)	-0.0003 (-0.12)	-0.0091*** (-3.28)
French	0.1186*** (13.96)	0.1087*** (20.08)	0.0294*** (6.33)	0.0493*** (11.82)
Bachelor's	0.0224*** (3.94)	0.0299*** (4.29)	0.0036 (1.15)	0.0173*** (3.87)
Master's/PhD	-0.0102* (-1.72)	-0.0093 (-1.50)	-0.0082*** (-2.78)	0.0128*** (3.74)
Expert	0.0181*** (3.81)	0.0255*** (5.34)	0.0000 (0.02)	-0.0119*** (-3.64)
Serial	-0.0495*** (-7.88)	-0.0319*** (-5.80)	-0.0054* (-1.80)	-0.0715*** (-21.57)
Incorporated	0.0509*** (5.12)	0.1366*** (14.11)	-0.0414*** (-7.94)	-0.0475*** (-12.76)
Tangible/total assets	0.1811*** (15.03)	0.1784*** (16.05)	0.0518*** (7.29)	0.0732*** (9.43)
Ln(total assets)	0.0203*** (8.54)	0.0266*** (9.27)	0.0047*** (4.46)	-0.0051*** (-5.10)
Sector $\times$ cohort FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
$R^2$	0.162	0.134	0.055	0.241
N	92338	92338	92338	92338

**Table 1.4. The Effects of Gender Stereotypes on External Equity and Other Financing Sources**

*Source:* SINE surveys and tax files. *Sample:* Repeated cross-section of new firms founded in 2002, 2006, 2010, and 2014. The table reports linear probability model estimates and analyzes the effect of gender stereotypes on the use of different financing sources. The dependent variables are as follows: a dummy variable that is equal to 1 if the start-up receives VC or other external equity financing (panel A, columns 1-4); a dummy variable that is equal to 1 if the start-up receives any type of external financing (panel B, column 1); a dummy variable that is equal to 1 if the start-up receives a bank loan (panel B, column 2); a dummy variable that is equal to 1 if the entrepreneur has received a personal bank loans (panel B, column 3); a dummy variable equal to 1 if the start-up receives an equity subsidy (panel B, column 4). *Female-dominated sector* (Entrepreneurs) in column 1 is a dummy variable that takes a value of 1 if more than 50% of new start-ups within a 4-digit French SIC sector are female-founded. Other measures of *Female-dominated sector* also capture within sector female representativeness and are based on the % of new female-founded businesses (panel A, column 2), % of female small business owners (panel A, column 3) and % of female CEOs (panel A, column 4). Control variables at the 4-digit French SIC level are the within-sector percentage of female-founded start-ups, Herfindahl index and the logarithm of total sector sales. Human capital controls include entrepreneur's age and French citizenship, education and experience dummy variables. Start-up controls include an incorporation status dummy variable, the ratio of tangible assets and the logarithm of firm's total assets. All models include 4-digit French SIC sector, cohort-year and county fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at the 10, 5, and 1% levels, respectively.

Panel A. External Equity				
Dependent variable:	External Equity			
	Entrepreneurs	New business owners	All business owners	CEOs
Measure of female-dominated sector:	(1)	(2)	(3)	(4)
Female	-0.0069*** (-5.45)	-0.0081*** (-6.33)	-0.0073*** (-5.72)	-0.0077*** (-5.28)
Female-dominated sector	-0.0000 (-0.00)	-0.0005 (-0.14)	-0.0003 (-0.08)	-0.0015 (-0.38)
F × F-dominated sector	0.0080*** (3.23)	0.0105*** (3.93)	0.0078*** (2.71)	0.0052** (2.04)
Human capital controls	Yes	Yes	Yes	Yes
Start-up controls	Yes	Yes	Yes	Yes
Additional sector controls	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Cohort (cohort) FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
$R^2$	0.026	0.026	0.026	0.026
N	92529	92356	92439	92529

**The Effects of Gender Stereotypes on External Equity and Other  
Financing Sources**  
*(Continued)*

Panel B. Alternative Financing Sources

Dependent variable:	External financing (1)	Bank loans (2)	Personal loans (3)	Subsidies (4)
Female	0.0023 (0.37)	-0.0039 (-0.83)	0.0066** (2.23)	-0.0029 (-0.73)
Female-dominated sector	0.0424 (1.44)	0.0451* (1.87)	0.0120 (0.98)	0.0203 (0.89)
F × F-dominated sector	0.0261 (1.51)	0.0069 (0.44)	0.0018 (0.21)	0.0350*** (2.73)
Human capital controls	Yes	Yes	Yes	Yes
Additional sector controls	Yes	Yes	Yes	Yes
Start-up controls	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
$R^2$	0.147	0.121	0.041	0.221
N	92529	92529	92529	92529

**Table 1.5. Gender Stereotypes and Future Corporate Performance**

*Source:* SINE surveys, tax files and employers' payrolls. *Sample:* Panel (columns 1-4) and repeated cross sections (columns 5-8) of successfully equity-funded start-ups founded in 2002, 2006, 2010, and 2014. The table reports OLS estimates and examines how funded minority entrepreneurs perform relative to entrepreneurs from the dominant gender group within and across sectors. Panel A examines performance of female entrepreneurs in male dominated sectors. Panel B examines performance of male entrepreneurs in female-dominated sectors. The *Female* gender is interacted with *Male-dominated sector* which is a dummy variable that takes a value of 1 if more than 50% of firms within a 4-digit French SIC are male-founded. The dependent variables are measures of future corporate development, performance and risk, and include the *Number of employees* (column 1), the ratio of *Sales over assets* (column 2), *% foreign sales* (column 3), the *ROA* (column 4), from one year after creation up to five years onwards. Other measures of performance include dummy variables *Survival 5 years* and *Survival 3 years* that takes the value one if the firm has survived at least 5 or 3 years, respectively (columns 5-6). Measures of corporate risk include firms' volatility of employment and sales, respectively, within firm from 1 year after creation up to 5 years (columns 7 and 8). All regressions include human capital and start-up fixed effects, as well as county, 4-digit French SIC sector and cohort-year fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Panel A. Female entrepreneurs in male-dominated sectors								
Dependent variables:	N. Employees (1)	Sales over assets (2)	% foreign sales (3)	ROA (4)	Survival 5 years (5)	Survival 3 years (6)	Vol. em- ployment (7)	Vol. sales (8)
Female	-3.1679** (-2.05)	-0.5700*** (-2.70)	-0.0207** (-2.42)	0.0505*** (2.63)	0.0032 (0.05)	0.0555 (0.94)	-38.5075 (-1.61)	-38.5075 (-1.61)
Male-dominated sector	-1.7976 (-1.40)	-0.5967 (-1.57)	-0.0157 (-1.35)	0.0300 (1.26)	-0.0658 (-0.80)	-0.0607 (-0.81)	32.6155 (0.69)	32.6155 (0.69)
Female × M-dominated sector	3.6962** (2.26)	0.7132*** (2.77)	0.0271*** (2.85)	-0.0644*** (-3.00)	-0.0539 (-0.73)	-0.0882 (-1.38)	33.7830 (0.92)	33.7830 (0.92)
Human capital controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.257	0.114	0.242	0.120	0.160	0.158	0.200	0.200
N	8708	8885	8893	8752	2705	2705	2334	2334

Panel B. Male entrepreneurs in female-dominated sectors								
Dependent variables:	N. Employees (1)	Sales over assets (2)	% foreign sales (3)	ROA (4)	Survival 5 years (5)	Survival 3 years (6)	Vol. em- ployment (7)	Vol. sales (8)
Male	-0.5284 (-1.07)	-0.1433 (-0.85)	-0.0064 (-1.33)	0.0138 (1.39)	0.0508* (1.82)	0.0327 (1.29)	4.7245 (0.16)	4.7245 (0.16)
Female-dominated sector	-1.8987 (-1.54)	-0.1166 (-0.36)	-0.0115 (-0.81)	0.0344 (1.44)	0.1198 (1.31)	0.1489* (1.83)	-66.3985 (-1.23)	-66.3985 (-1.23)
Male × F-dominated sector	3.6962** (2.26)	0.7132*** (2.77)	0.0271*** (2.85)	-0.0644*** (-3.00)	-0.0539 (-0.73)	-0.0882 (-1.38)	33.7830 (0.92)	33.7830 (0.92)
Human capital controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.257	0.114	0.242	0.120	0.160	0.158	0.200	0.200
N	8708	8885	8893	8752	2705	2705	2334	2334

**Table 1.6. Self-selection into Sectors**

*Source:* SINE surveys. *Sample:* Repeated cross-section of new firms founded in 2002, 2006, 2010 and 2014. The table reports linear probability model estimates and analyzes the choice of entrepreneurs between female- versus male-dominated sectors. Panel A includes all start-ups. Panel B includes only start-ups that receive VC or other external equity financing. The dependent variable is the dummy variable *Female-dominated sector* that is equal to 1 if at least 50% of new start-ups within a 4-digit French SIC sector are female-founded. The independent variables are the start-up founder's personal traits and motivation items interacted with the founder's gender. All models include human capital controls and cohort-year fixed effects. Human capital controls include founder's age, French citizen, education and experience dummy variables. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Panel A. All start-ups								
Dependent variable:	1{Female-dominated sector}							
Item:	Elite engineer school (1)	Serial (2)	Optimistic (3)	High- growth oriented (4)	New idea (5)	Opportunity (6)	Taste (7)	Independence (8)
Female × Item	-0.1164* (-1.74)	-0.0361** (-2.07)	-0.1202*** (-3.65)	-0.0759*** (-3.40)	-0.0432 (-1.32)	-0.0258** (-2.23)	0.0054 (0.35)	0.0653* (1.88)
Item	0.0017 (0.18)	0.0074 (1.01)	-0.0015 (-0.13)	0.0011 (0.18)	0.0112 (1.18)	0.0060 (1.19)	-0.0006 (-0.11)	-0.0021 (-0.44)
Female	0.2182*** (3.07)	0.2246*** (2.95)	0.2542*** (3.23)	0.2439*** (3.14)	0.2229*** (2.89)	0.2213*** (3.04)	0.2135*** (3.07)	0.1770*** (3.31)
Other human capital controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year (cohort) FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.197	0.197	0.224	0.208	0.197	0.196	0.196	0.199
N	92446	92446	55404	80462	92439	92439	92441	92444
Panel B. Equity-backed start-ups								
Dependent variable:	1{Female-dominated sector}							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female × Item	-0.0403 (-0.60)	0.0140 (0.24)	-0.1131 (-1.39)	-0.0813 (-1.59)	-0.0172 (-0.30)	-0.0326 (-0.83)	-0.0308 (-0.63)	0.0438 (1.02)
Item	0.0509** (2.09)	0.0071 (0.50)	0.0044 (0.29)	-0.0013 (-0.11)	0.0167 (0.87)	0.0134 (1.13)	-0.0111 (-1.12)	-0.0123 (-1.03)
Female	0.1896*** (3.28)	0.1813*** (3.26)	0.2262*** (2.94)	0.2287*** (3.47)	0.1894*** (3.15)	0.1949*** (3.24)	0.1991*** (3.54)	0.1645*** (2.97)
Other human capital controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year (cohort) FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.143	0.144	0.156	0.163	0.144	0.144	0.145	0.145
N	2044	2044	1337	1550	2043	2043	2043	2044

**Table 1.7. Gender Stereotypes and Human Capital**

*Source:* SINE surveys and tax files. *Sample:* Repeated cross-section of new firms founded in 2002, 2006, 2010, and 2014. The table reports linear probability model estimates and analyzes the joint effects of human capital and gender stereotypes on the use of external equity financing. The dependent variable is a dummy variable that is equal to 1 if the start-up receives VC or other external equity financing. *Female-dominated sector* is a dummy variable that takes a value of 1 if more than 50% of new start-ups within a 4-digit French SIC are female-founded. This variable is interacted with various motivation items: *Master's/PhD* (column 1) is equal to 1 if the entrepreneur has a master's or higher degree; *Elite engineering school* is a dummy variable that takes a value of 1 if the entrepreneur graduated from a top engineering school in France (column 2); *Expert* (column 3), is equal to 1 if the entrepreneur has already worked in the industry prior to the entry in entrepreneurship; and *Serial* (column 4) if the entrepreneur has already founded a company in the past. All models include other human capital variable that are not interacted, start-ups controls and sector controls. They also include county, 4-digit French SIC sector and cohort-year fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Dependent variable: Item:	External equity			
	Master's/ PhD (1)	Elite engineering school (2)	Expert (3)	Serial (4)
Female	-0.0053*** (-3.83)	-0.0080*** (-6.32)	-0.0073*** (-4.18)	-0.0036** (-2.45)
Female-dominated sector	-0.0003 (-0.07)	-0.0027 (-0.67)	-0.0024 (-0.45)	0.0004 (0.12)
F × Female-dominated sector	0.0088*** (2.80)	0.0103*** (4.08)	0.0116** (2.38)	0.0074*** (2.59)
Item	0.0106*** (5.72)	0.0098** (2.50)	-0.0003 (-0.20)	0.0161*** (9.89)
F × Item	-0.0063** (-2.56)	0.0037 (0.40)	0.0007 (0.30)	-0.0132*** (-4.36)
F-dominated sector × Item	0.0003 (0.04)	0.0338** (2.17)	0.0035 (0.62)	-0.0018 (-0.30)
F × F-dominated sector × Item	-0.0009 (-0.14)	-0.0216 (-0.80)	-0.0051 (-0.81)	0.0035 (0.41)
Additional sector controls	Yes	Yes	Yes	Yes
Other human capital controls	Yes	Yes	Yes	Yes
Start-up controls	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Year (cohort) FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.025	0.026	0.026	0.026
N	80503	92520	92524	92522

**Table 1.8. Gender Stereotypes, Team and Family Structure**

*Source:* SINE surveys and tax files. *Sample:* Repeated cross-section of new firms founded in 2002, 2006, 2010, and 2014. The table reports linear probability model estimates and analyzes the joint effects of starting as a team and gender stereotypes on the use of external equity financing. The dependent variable is a dummy variable that is equal to 1 if the receives VC or other external equity financing. The effect of gender stereotypes is tested on different sub-samples: a subsample of start-ups started by stand-alone entrepreneurs (column 1), a subsample of start-ups founded by teams (column 2), a subsample of teams formed by spouses (column 2), a subsample of teams formed by family members (column 4), a subsample of teams formed by associates (column 5), a subsample of married entrepreneurs (column 6), and a subsample of single entrepreneurs (column 7). All models include human capital controls, start-ups controls and sector controls. They also include county, 4-digit French SIC sector and cohort-year fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Dependent variable: Sub-sample:	External equity						
	Alone	Team	Team			Married	Single
	(1)	(2)	Spouse	Family	Associate	(6)	(7)
			(3)	(4)	(5)		
Female	-0.0048*** (-3.74)	-0.0145*** (-5.36)	-0.0056 (-1.58)	-0.0120** (-2.43)	-0.0167*** (-4.01)	-0.0046*** (-2.90)	-0.0132*** (-6.68)
Female-dominated sector	-0.0002 (-0.04)	-0.0116 (-1.37)	-0.0072 (-0.73)	-0.0242* (-1.86)	-0.0066 (-0.54)	-0.0030 (-0.49)	-0.0041 (-0.60)
Female × Female-dominated sector	0.0078*** (3.02)	0.0113* (1.82)	0.0068 (0.88)	0.0115 (1.26)	0.0025 (0.32)	0.0032 (0.97)	0.0201*** (4.07)
Sector controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Human capital controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Start-up controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year (cohort) FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.025	0.049	0.051	0.095	0.070	0.029	0.043
N	65878	24206	11940	6422	13684	69626	22838

**Table 1.9. Gender Stereotypes, Initial Motivations and Optimism**

*Source:* SINE surveys and tax files. *Sample:* Repeated cross-section of new firms founded in 2002, 2006, 2010, and 2014. The table reports linear probability model estimates and analyzes the joint effects of initial motivations and gender stereotypes on the use of external equity financing. The dependent variable is a dummy variable that is equal to 1 if the start-up receives VC or other external equity financing. *Female-dominated sector* is a dummy variable that takes a value of 1 if more than 50% of new start-ups within a 4-digit French SIC are female-founded. This variable is interacted with various motivation items: *High-growth oriented* (column 1) is equal to 1 if the entrepreneur intends to develop the company and 0 if he intends to become self-employed; *New idea* (column 2), *Taste* (column 3), *Opportunity* (column 4), and *Independence* (column 5) stem from the question "What are your three main motivations?", and correspond to "a new idea of product, service, or market", "the taste for entrepreneurship or new challenges", "an opportunity to create a start-up", and "the desire to be independent", respectively. *Optimism at start* (column 6) is the difference between initial hiring expectations and subsequent realizations (Landier and Thesmar, 2009). Expectation is equal to 1 when the entrepreneur answers "Yes" to the question "Do you plan to hire over the next 12 months?" Realization is equal to 1 if the firm labor force increases by at least two employees in the year after the creation. All models include human capital controls, start-ups controls and sector controls. They also include county, 4-digit French SIC sector and cohort-year fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Dependent variable: Item:	External equity					
	High-growth oriented (1)	Ex ante Motivations				Optimism at start (6)
		New idea (2)	Taste (3)	Opportunity (4)	Independence (5)	
Female	-0.0037*** (-2.95)	-0.0060*** (-4.48)	-0.0045*** (-3.07)	-0.0070*** (-5.39)	-0.0108*** (-5.17)	-0.0062*** (-3.77)
Female-dominated sector	-0.0035 (-0.84)	-0.0014 (-0.28)	0.0041 (0.80)	-0.0004 (-0.09)	0.0032 (0.49)	0.0030 (0.53)
F × Female-dominated sector	0.0090*** (2.99)	0.0095*** (2.95)	0.0062* (1.67)	0.0091*** (3.93)	0.0062 (1.09)	0.0111*** (2.90)
Item	0.0119*** (7.45)	0.0086*** (4.26)	0.0056*** (4.27)	0.0051*** (3.32)	-0.0101*** (-8.35)	0.0083*** (3.83)
F × Item	-0.0091*** (-3.65)	-0.0049 (-1.33)	-0.0052** (-2.32)	-0.0001 (-0.02)	0.0062** (2.36)	-0.0076* (-1.94)
F-dominated sector × Item	-0.0035 (-0.70)	0.0063 (0.58)	-0.0098*** (-2.75)	0.0009 (0.16)	-0.0062 (-1.13)	0.0020 (0.29)
F × F-dominated sector × Item	-0.0001 (-0.02)	-0.0051 (-0.51)	0.0048 (0.97)	-0.0022 (-0.28)	0.0045 (0.68)	-0.0073 (-0.62)
Additional sector controls	Yes	Yes	Yes	Yes	Yes	Yes
Human capital controls	Yes	Yes	Yes	Yes	Yes	Yes
Start-up controls	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Year (cohort) FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.025	0.026	0.026	0.026	0.027	0.032
N	80505	92520	92524	92522	92527	61768

**Table 1.10. The Effects of Age Stereotypes on External Financing**

*Source:* SINE surveys and tax files. *Sample:* Repeated cross-section of new firms founded in 2002, 2006, 2010, and 2014. The table reports linear probability model estimates and analyzes the effect of age stereotypes on the use of financing sources. The dependent variables are a dummy variable that is equal to 1 if the start-up receives VC or other external equity financing (column 1), a dummy variable that is equal to 1 if the start-up receives bank loans (column 2), a dummy variable that is equal to 1 if the entrepreneur has received a personal bank loan (column 3), and a dummy variable that is equal to 1 if the entrepreneur receives an equity subsidy (column 4).  $Age \geq 50$  is a dummy variable that takes the value of 1 if the entrepreneur is 50 years old or older. *Young-CEO sector* is a dummy variable that takes the value of 1 if the median CEO age within a 4-digit French SIC sector is younger than 40 years old. Control variables at the 4-digit French SIC level are the Herfindahl index, the logarithm of total sector sales and the within sector percentage of female-founded stat-ups. Human capital controls include French citizenship, female-gender, education and experience dummy variables. Start-up controls include an incorporation status dummy variable, the ratio of tangible assets and the logarithm of firm's total assets. All regressions include county, 4-digit French SIC sector and cohort-year fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Dependent variable:	External equity (1)	Bank loans (2)	Personal loans (3)	Subsidies (4)
Age $\geq$ 50 years old	0.0015 (0.95)	-0.0694*** (-15.96)	-0.0061** (-2.24)	-0.0228*** (-6.41)
Young-CEO sector	0.0182** (2.48)	-0.0445 (-0.62)	0.0010 (0.05)	0.0619 (1.22)
Age $\geq$ 50 x Young-CEO sector	-0.0358*** (-3.11)	0.0295 (0.36)	0.0578 (1.33)	0.0098 (0.44)
Additional sector controls	Yes	Yes	Yes	Yes
Human capital controls	Yes	Yes	Yes	Yes
Start-up controls	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Year (cohort) FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
$R^2$	0.026	0.122	0.041	0.221
N	92524	92524	92524	92524

**Table 1.11. Homophily, Gender of the Investor and Gender of the Entrepreneur**

The table reports linear probability model estimates and tests whether investors are more likely to invest in start-ups founded by entrepreneurs of the same gender. Note that the test of the homophily hypothesis is conditional on receiving any kind of external equity as the gender of the investors is known only if the firm received external equity. The models do not include fixed effects to live enough degrees of freedom. Robust standard errors are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively. *Source:* VentureXpert, tax files and employer payrolls databases. *Sample:* Equity backed firms available in VentureXpert from 2002 to 2016 that could be linked to the tax files and employer payrolls databases.

dependent variable	Female (1)	Venture capital (2)	Private equity (3)	Angel (4)	CVC (5)
Constant	0.128*** (12.97)	0.113*** (8.20)	0.129*** (10.03)	0.145*** (5.36)	0.167** (2.56)
Female VC	0.029 (0.81)	0.007 (0.15)	0.021 (0.44)	0.135 (1.42)	-0.167** (-2.56)
Female-dominated sector	0.176*** (4.68)	0.202*** (3.11)	0.156*** (3.44)	0.127 (0.92)	-0.167** (-2.56)
Female VC times Female-dominated sector	-0.103 (-0.81)	-0.322*** (-4.04)	-0.083 (-0.54)	0.593*** (3.58)	0.167** (2.56)
$R^2$	0.024	0.029	0.021	0.062	0.032
N	1430	641	861	210	43

**Table 1.12. Homophily, Gender of the Investor and Future Corporate Performance**

*Source:* VentureXpert, tax files and employer payrolls. *Sample:* PE- and VC-backed firms available in VentureXpert from 2002 to 2016 that could be linked to the tax files and employer payrolls databases. The table reports OLS estimates and examines the joint effects of receiving external equity financing and of the investor's gender on corporate performance and development from one year after creation to five years onward. Dependent variables are the start-ups' *Number of employees*, *Logarithm (1+ Number of employees)*, *Sales over assets* and the *ROA*. *Female entrepreneur* is a dummy variable that is equal to 1 if the start-up is female-led. *Female VC* is a dummy variable that is equal to 1 if the PE or VC investment firm is led by a female CEO. The models include year, county and sector fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Dependent variables:	N. employees (1)	Ln(1+employees) (2)	Sales over assets (3)	ROA (4)
Female	-4.824 (-0.88)	-0.319*** (-6.35)	-0.010 (-0.27)	-0.001 (-0.06)
Post	15.170*** (5.18)	0.492*** (17.19)	-0.036* (-1.77)	-0.060*** (-8.29)
Female × Post	1.672 (0.21)	0.094 (1.30)	-0.010 (-0.22)	0.009 (0.60)
Female VC	16.774*** (3.02)	0.148* (1.95)	0.082* (1.67)	0.008 (0.54)
Female × Female VC	14.035 (0.89)	0.156 (0.73)	-0.040 (-0.30)	0.021 (0.53)
Post × Female VC	-12.484* (-1.88)	0.040 (0.45)	0.034 (0.56)	0.046** (2.22)
Female × Post × Female VC	4.785 (0.24)	-0.133 (-0.44)	-0.164 (-1.01)	0.021 (0.35)
Sector FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
$R^2$	0.655	0.599	0.603	0.347
N	10367	10229	9664	9319

**Table 1.13: The Effect of Experience on Gender Stereotypes**

The table reports linear probability model estimates and tests the effect of experience on stereotypical investment decision. It tests the investors' likelihood to invest in a female-led start-up when the investment firm has already invested in female-led start-ups in the past. The dependent variable is the entrepreneur's gender, *Female*. The main independent variables are the dummy variable Female-dominated sector and *Experience*, that reflects the investor's experience with female-led start-ups. *Experience* is a dummy variable that takes the value 1 if the investor has at least one female-led start-up in portfolio the year before investing in the start-up (Panel A), or it is the percentage of female-led start-ups in portfolio the year before investing in the start-up (Panel B). Note that only firms that received external equity investment are included in samples. Across the different columns the relationship between entrepreneur's gender and investor's experience is estimated in different sub-sample, depending on the investment type. Robust standard errors are reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at the 10, 5, and 1% levels, respectively. *Source*: VentureXpert. PE- and VC-backed deals available in VentureXpert from 2002 to 2016 that could be linked to the tax files and employer payrolls databases.

Panel A: Investment portfolio with at least one female-led start-up

	External equity (1)	Venture capital (2)	Private equity (3)	Angel (4)	CVC (5)
Experience	0.035** (2.01)	0.021 (0.89)	0.040* (1.70)	-0.060 (-1.02)	0.062 (0.94)
Female-dominated sector	0.064 (1.50)	0.097 (1.30)	0.020 (0.42)	-0.064 (-0.45)	-0.071* (-1.76)
Experience $\times$ F-dominated sector	0.108* (1.79)	0.029 (0.28)	0.155** (2.23)	0.251 (1.17)	-0.062 (-0.94)
Constant	0.108*** (8.06)	0.103*** (5.96)	0.107*** (5.69)	0.207*** (3.86)	0.071* (1.76)
$R^2$	0.020	0.012	0.022	0.011	0.021
N	1790	773	1079	272	96

Panel B: Percentage of female entrepreneurs in the investment portfolio

	External equity (1)	Venture capital (2)	Private equity (3)	Angel (4)	CVC (5)
Experience	0.140* (1.84)	0.068 (0.79)	0.189 (1.50)	0.258 (1.44)	-0.158 (-0.19)
Female-dominated sector	0.115*** (3.16)	0.091 (1.59)	0.108** (2.54)	-0.054 (-0.54)	-0.117*** (-2.66)
Experience $\times$ F-dominated sector	0.238 (0.84)	0.084 (0.24)	0.248 (0.78)	2.620*** (3.43)	0.158 (0.19)
Constant	0.125*** (13.03)	0.115*** (8.46)	0.126*** (9.99)	0.134*** (5.07)	0.117*** (2.66)
$R^2$	0.018	0.008	0.019	0.041	0.013
N	1709	724	1044	269	89

## 1.10. Description of variables

Variable	Description
<u>Dependent variables (Source: SINE)</u>	
External equity	Dummy variable that is equal to 1 if the start-up receives venture capital or other equity financing at creation and zero otherwise.
Venture capital	Dummy variable that is equal to 1 if the start-up receives VC financing at creation and zero otherwise. It is available only in 2002, 2010, 2014.
External financing	Dummy variable that is equal to 1 if the start-up receives any sources of external financing and zero otherwise.
Bank loan	Dummy variable that is equal to 1 if the start-up receives a bank loan granted to the start-up at creation and zero otherwise.
Personal loan	Dummy variable that is equal to 1 if the start-up receives a bank loan granted to the founder at creation and zero otherwise.
Subsidies	Dummy variable that is equal to 1 if the start-up receives an equity grant at creation and zero otherwise.
Personal resources	Dummy variable that is equal to 1 if the founder invests personal resources at creation and zero otherwise.
<u>Dependent variables (Source: VentureXpert)</u>	
External equity	Dummy variable that is equal to 1 if the company receives VC, other private equity, corporate venture capital, or angel investment financing and zero otherwise.
Venture capital	Dummy variable that is equal to 1 if the company receives VC financing and zero otherwise.
Private equity	Dummy variable that is equal to 1 if the company receives PE financing and zero otherwise.
CVC	Dummy variable that is equal to 1 if the company receives corporate venture capital financing and zero otherwise.
Angel	Dummy variable that is equal to 1 if the company receives angel investment financing and zero otherwise.
<u>Key independent variables (Source: SINE)</u>	
Female	Dummy variable that is equal to 1 if the start-up is founded by a female entrepreneur and zero if founded by a male entrepreneur.
Age $\geq 40$	Dummy variable that is equal to 1 if the entrepreneur is 40 years old or older at creation.
Age $\geq 50$	Dummy variable that is equal to 1 if the entrepreneur is 50 years old or older at creation.
French	Dummy variable that is equal to 1 if the entrepreneur is a French citizen and zero otherwise.
High school	Dummy variable that is equal to 1 if the entrepreneur's highest degree is a high school diploma (Baccalauréat) and zero otherwise.
Bachelor's	Dummy variable that is equal to 1 if the entrepreneur's highest diploma is a three-year bachelor's degree (Licence) and zero otherwise.
Master's/PhD	Dummy variable that is equal to 1 if the entrepreneur has at least a five-year master's degree, including engineering, JD, MD, and PhD degrees (Master, Grande école, Doctorat), and zero otherwise.
Elite engineering school	Dummy variable that is equal to 1 if the entrepreneur has graduated from a top engineering school (Grande école d'ingénieur, e.g., Ecole Polytechnique, Centrale, Mines among others), and zero otherwise.
Expert	Dummy variable that is equal to 1 if the entrepreneur has at least three years of prior work experience in the sector in which the start-up is incorporated and zero otherwise.
Serial	Dummy variable that is equal to 1 if the entrepreneur has already founded a start-up and zero otherwise.

*Continued on next page*

Variable	Description
Optimism at start	Dummy variable that corresponds to the difference between initial hiring expectations and subsequent realizations (Thesmar and Landier, 2009). Expectation is equal to 1 when the entrepreneur answers "Yes" to the question "Do you plan to hire over the next 12 months?" and zero otherwise. Realization is equal to 1 if the firm labor force increases by at least two employees in the year after creation and zero otherwise. Note that Optimism at start takes the value of zero if Realization > Expectation. Optimism is also computed at periods t+3 and t+5. Sources: SINE first period and employer payrolls
High-growth oriented	Dummy variable that stems from the question "What is your main objective?" and takes a value of 1 if the entrepreneur answers "to develop the company" but zero she answers "mainly to create my own job".
Motivation items stem from the question "What are your three main motivations?":	
New idea	Dummy variable that is equal to 1 if entrepreneur ticks the box "a new idea for a product, service, or market" and zero otherwise.
Taste	Dummy variable that is equal to 1 if entrepreneur ticks the box "taste for entrepreneurship or new challenges" and zero otherwise.
Opportunity	Dummy variable that is equal to 1 if entrepreneur ticks the box "an opportunity to create a start-up" and zero otherwise.
Independence	Dummy variable that is equal to 1 if the entrepreneur ticks the box "desire to be independent" and zero otherwise.
Founding Team:	
Alone	Dummy variable that is equal to 1 if the entrepreneur indicates having started on her own and zero otherwise.
Spouse	Dummy variable that is equal to 1 if the entrepreneur indicates having started the company with her spouse and zero otherwise.
Family	Dummy variable that is equal to 1 if the entrepreneur indicates having started with a sibling, a relative or a friend, and zero otherwise.
Associate	Dummy variable that is equal to 1 if the entrepreneur indicates having started with a professional partner or an associate, and zero otherwise.
Married	Dummy that is equal to 1 if the entrepreneur is married or in a spousal relationship, and zero otherwise.
Children	Dummy variable that is equal to 1 if the entrepreneur has at least one child at the start-up creation date and zero otherwise.
<u>Female-dominated sectors and sectoral Characteristics</u>	
Female-dominated sector measures (at the 4-digit French SIC level):	
Entrepreneurs	Dummy variable that is equal to 1 if more than 50% of start-ups created within a sector are founded by a female entrepreneur and zero if less.
CEOs	Dummy variable that is equal to 1 if more than 50% of firms within a sector are led by a female CEO and zero if less. CEOs are identified with a 4-digit occupation code, or if not available, the CEO is assumed to be the highest paid employee. <i>Source</i> : Employer payrolls
Business Owners	Dummy variable that is equal to 1 if more than 50% of non-incorporated firms within a sector are owned by women and zero if less. <i>Source</i> : Firm registries
New Business Owners	Dummy variable that is equal to 1 if more than 50% of new firms within a sector are owned by women and zero if less. <i>Source</i> : Firm registries
% Female-founded	Percentage of female-founded start-ups within a sector. <i>Source</i> : SINE
Herfindahl	Herfindahl index based on sales. <i>Source</i> : Tax files
Log(total sector sales)	Logarithm of the sum of sales realized in a sector. <i>Source</i> : Tax files
Young sector	Dummy variable that is equal to 1 if the median CEO age within a 4-digit French SIC sector is lower than 40 years old and zero if higher. <i>Source</i> : SINE
Balance sheet and performance variables (Sources: Tax files & Employer payrolls)	

*Continued on next page*

Variable	Description
Tangible/ total assets	Tangible ratio is the sum of tangible assets divided by the balance sheet total assets.
Log(total assets)	Logarithm of the total assets on the balance sheet.
Survival 5 years	Dummy variable that is equal to 1 if the start-up survives five years after creation and zero otherwise.
Employment size at start	Number of employees at the end of the first year. Employment size is also created at periods three years after creation (t+3) and five years after creation (t+5).
N. employees	Number of employees.
Sales over assets	Total sales reported in P&L statements divided by total assets reported in balance sheets.
% foreign sales	Percentage of total sales realized abroad.
ROA	Returns on assets is the net income divided by the balance sheet total assets.
Investment firms characteristics (Sources: VentureXpert & Employer payrolls)	
Female VC	Dummy variable that is equal to 1 if the PE or VC investment firm is led by a female CEO (general partner). PE and VC firms are merged to the employer payrolls database using a Python webcrawler. Female CEOs are identified using an occupation code, or if not available, as the highest paid employee in employer payrolls database.
VC board has one women	Dummy variable that is equal to 1 if at least one women is in the top 5 of the most paid employees at the PE or VC investment firm, and zero otherwise.
Experience	

## 1.11. Appendix

### 1.11.1. Additional Tables

**Table 1.A1. Percentage of Female-founded Start-ups by Sector**

*Source:* SINE survey. Panel A reports the top and bottom five 4-digit French SIC sectors by share of within-sector new female-founded start-ups. Sectors with less than 30 start-ups are excluded. Panel B reports the ranking at the 1-digit French SIC.

Panel A. Top 10 and bottom 5 at the 4-digit French SIC level			
Rank	Sector (4-digit French SIC)	% Female	# Start-ups
1	Hairdressing and other beauty treatment	0.774	5,627
2	Manufacture of imitation jewelry and related articles	0.772	228
3	Other human health activities	0.707	5368
4	Manufacture of ceramic household and ornamental articles	0.702	114
5	Translation and interpretation activities	0.684	329
6	Physical well-being activities	0.678	541
7	Retail sale of cosmetic and toilet articles in specialized stores	0.670	218
8	Manufacture of other textiles n.e.c.	0.660	53
9	Retail sale of flowers, plants, seeds, fertilizers, pets and pet food	0.643	737
10	Retail sale of textiles in specialized stores	0.643	235
...			
280	Electrical installation	0.053	2,815
281	Repair of electrical equipment	0.042	71
282	Forging, pressing, stamping and roll-forming of metal; powder metallurgy	0.037	82
283	Manufacture of locks and hinges	0.032	31
284	Manufacture of other general-purpose machinery n.e.c.	0.027	37
Panel B. Top sectors at the 1-digit French SIC level			
Rank	Sector (1-digit French SIC)	% Female	# Start-ups
1	Other service activities	0.620	10,420
2	Human health and social work activities	0.619	7,984
3	Education	0.396	3541
4	Accommodation and food service activities	0.359	17,342
5	Arts, entertainment and recreation	0.358	3,560
6	Wholesale and retail trade, repair of motor vehicles and motorcycles	0.329	31,710
7	Real estate activities	0.321	6,358
8	Professional, scientific and technical activities	0.306	14,620
9	Administrative and support service activities	0.287	9,133
10	Manufacturing	0.256	11601
11	Financial and insurance activities	0.251	2,746
12	Information and communication	0.186	5,754
13	Water supply, sewerage, waste management and remediation activities	0.169	5,90
14	Transportation and storage	0.168	5,851
15	Electricity, gas, steam and air conditioning supply	0.142	1,184
16	Mining and quarrying	0.135	74
17	Agriculture, forestry and fishing	0.130	23
18	Construction	0.084	25,599

**Table 1.A2. Top 20 VC and External Equity-intensive Sectors**

*Source:* SINE survey. *Sample:* New firms founded in 2002, 2006, 2010, and 2014. The table reports the top 20 sectors which include the highest number of start-ups that receive VC and other external equity financing. The table also reports the number those start-ups that are female-founded, as well as the share of female-founded start-ups within the sector.

Rank	Sector (4-digit French SIC)	# Backed	# Backed female-led	% Female	# Start-ups
1	Other specialized construction activities	38	2	0.090	5,629
2	Restaurants and mobile food service activities	31	4	0.333	7,627
3	Joinery installation	26	2	0.062	2,576
4	Painting and glazing	22	0	0.092	2,761
5	Freight transport by road	18	4	0.165	2,456
6	Electrical installation	17	0	0.045	2,641
7	Computer programming activities	16	0	0.156	1,313
8	Plumbing, heat and air-conditioning installation	15	1	0.064	2,405
9	Construction of buildings	13	3	0.110	1677
10	Maintenance and repair of motor vehicles	11	0	0.119	1,498
11	Business and other management consultancy activities	11	2	0.278	4,372
12	Real estate agencies	10	1	0.354	3,990
13	Roofing activities	10	0	0.068	1,201
14	Engineering activities and related technical consultancy	10	1	0.128	1,851
15	General cleaning of buildings	9	2	0.308	1,411
16	Other retail sale of new goods in specialized stores	9	3	0.376	1,674
17	Manufacture of bread,fresh pastry goods and cakes	9	6	0.258	1,088
18	Hairdressing and other beauty treatment	9	6	0.780	4,186
19	Sale of cars and light motor vehicles	9	1	0.161	1,190
20	Buying and selling of own real estate	8	0	0.205	1,391

**Table 1.A3. Gender Distribution by Male- and Female-dominated Sectors**

*Source:* SINE survey. *Sample:* New firms started in 2002, 2006, 2010, and 2014. The table reports the percentages of female- and male-founded start-ups incorporated in female- and male-dominated sectors. A female-dominated sector include at least 50% of female-founded start-ups within a 4-digit French SIC sector.

	Cohorts				Total
	2002	2006	2010	2014	
<u>Male-dominated sector</u>	11264	13757	30762	26379	24274
%	89	88	91	87	26
Male	9243	11131	23501	20445	6432
%	73	71	70	67	70
Female	2021	2626	7261	5934	17842
%	16	17	22	20	19
<u>Female-dominated sector</u>	1350	1963	2918	4053	10284
%	11	12	9	13	11
Male	475	678	1112	1454	3719
%	4	4	3	5	4
Female-dominated sector	875	1285	1806	2599	6565
%	7	8	5	9	7
Total	12614	15720	33680	30432	92446

**Table 1.A4. Switching Sectors**

*Source:* SINE survey. 4-digit French SIC sectors that switch from male-dominated to female dominated between the 2010 cohort and the 2014 cohort.

Sector (4-digit French SIC)	Cohort Year	% female 2010	% female 2014	N. start-ups
Manufacture of rusks and biscuits; manufacture of preserved pastry goods and cakes	2014	0.46	0.57	28
Manufacture of cocoa, chocolate and sugar confectionery	2014	0.25	0.50	12
Manufacture of other outerwear	2014	0.43	0.57	214
Manufacture of perfumes and toilet preparations	2014	0.42	0.55	11
Manufacture and processing of other glass, including technical glassware	2014	0.47	0.60	10
Manufacture of jewellery and related articles	2014	0.24	0.52	31
Other manufacturing n.e.c.	2014	0.38	0.54	65
Agents involved in the sale of timber and building materials	2014	0.13	0.50	14
Dispensing chemist in specialised stores	2014	0.42	0.56	36
Retail sale of watches and jewellery in specialised stores	2014	0.46	0.55	22
Retail sale of second-hand goods in stores	2014	0.45	0.51	39
Holiday and other short-stay accommodation	2014	0.38	0.50	160
Veterinary activities	2014	0.49	0.52	67
Cultural education	2014	0.38	0.67	42
General medical practice activities	2014	0.47	0.51	333
Dental practice activities	2014	0.45	0.57	217
Social work activities without accommodation for the elderly and disabled	2014	0.47	0.60	68
Artistic creation	2014	0.43	0.52	308
Washing and (dry-)cleaning of textile and fur products	2014	0.43	0.55	86
Other personal service activities n.e.c.	2014	0.45	0.51	349

**Table 1.A5. Characteristics of Male and Female-dominated Sectors**

*Source:* SINE survey, tax files and employer payrolls. *Sample:* Sectors identified as male- or female-dominated in 2002, 2006, 2010, and 2014. The table reports characteristics differences between male- and female-dominated sectors and start-ups characteristics differences within male- and female-dominated sectors. A female-dominated sector includes at least 50% of female-founded start-ups within a 4-digit French SIC sector. The mean and number of observations by gender group are reported as well as mean differences and t-statistics between the two groups.

	Male-dominated sector	Female-dominated sector	difference	t-stat		
	(1)	(2)	(1-2)			
N. start-ups	1522	124.89	279	94.95	29.94*	(1.70)
% Female-founded	1522	0.18	279	0.66	-0.48***	(-42.43)
Total sales	1506	19077561.80	278	17106525.56	1971036.24	(0.99)
Vol sales	1504	63658.22	278	36217.68	27440.54***	(3.20)
External finance dependence	1222	2.24e+09	207	-33.70	2.24e+09	(0.74)
N. backed firms within sector	1522	3.61	279	1.59	2.02***	(5.57)
N. VC-backed firms within sector	1522	0.67	279	0.32	0.35***	(3.48)
<u>Start-ups' characteristics:</u>						
Mean % tangible assets	1506	-2.52e+08	278	0.20	-2.52e+08	(-0.41)
Mean cash ratio	1506	3.72e+08	278	40916686.47	3.31e+08	(1.41)
Mean Leverage	1506	5.20e+10	278	2.58e+11	-2.06e+11	(-0.82)
Mean ROA	1506	-3.40e+10	278	-3.62e+11	3.28e+11	(0.91)
Mean sales	1506	9097.40	278	5618.11	3479.29***	(2.66)
Mean N. employees	1506	29.28	278	21.03	8.25***	(2.93)

**Table 1.A6. Characteristics of Equity-Backed Start-up Founders**

*Source:* SINE survey. *Sample:* Equity backed firms founded in 2002, 2006, 2010, and 2014. This table presents entrepreneurs' biographical characteristics (panel A), motivation and optimism items (panel B), composition of the entrepreneurial team (panel C), start-up's alternative financing sources at creation (panel D). The mean and number of observations by group of male- and female-founded start-ups are reported as well as mean differences and t-statistics between the two groups. Variable definitions and data sources are provided in the variable definition table.

Variables	All	Male		Female		difference	t-stat
	Mean	N	Mean	N	Mean		
Panel A. Founder biographical characteristics							
Bachelor's	0.13	2032	0.12	471	0.18	-0.06***	(-3.19)
Master's/PhD	0.31	2032	0.31	471	0.29	0.02	(0.67)
Elite engineering school	0.09	2032	0.10	471	0.07	0.03***	(2.61)
Expert	0.67	2032	0.69	471	0.57	0.11***	(4.49)
Serial	0.46	2032	0.49	471	0.31	0.18***	(7.51)
Panel B. Founder initial motivations and optimism							
High-growth oriented	0.47	1528	0.50	377	0.37	0.13***	(4.58)
Independence	0.47	2032	0.47	471	0.49	-0.03	(-1.08)
Taste	0.53	2031	0.54	471	0.48	0.06**	(2.37)
Add earnings	0.24	1764	0.25	428	0.22	0.03	(1.24)
Opportunity	0.26	2031	0.25	471	0.29	-0.04*	(-1.66)
New idea	0.23	2031	0.23	471	0.24	-0.01	(-0.27)
Optimism at start	0.46	1111	0.50	254	0.30	0.20***	(6.28)
Panel C. Team composition							
Alone	0.56	1987	0.55	454	0.60	-0.04*	(-1.74)
Spouse	0.11	1987	0.10	454	0.16	-0.06***	(-3.12)
Relatives	0.06	1987	0.06	454	0.06	0.01	(0.50)
Associates	0.28	1988	0.30	454	0.21	0.09***	(4.04)
Panel D. Start-up alternative financing sources							
Bank loans	0.44	2032	0.43	471	0.46	-0.03	(-1.17)
Personal loans	0.11	2032	0.11	471	0.11	-0.00	(-0.20)
Non-bank loans	0.11	2032	0.10	471	0.13	-0.03*	(-1.67)
Subsidies	0.15	2032	0.15	471	0.14	0.00	(0.16)

**Table 1.A7. Personal Traits of Start-up Founders by Gender and Gender-dominated Sector**

*Source:* SINE survey. *Sample:* New firms founded in 2002, 2006, 2010, and 2014. The table reports differences by founder's gender between those who founded a start-up in a male-dominated sector and those who founded a start-up in a female-dominated sector. A female-dominated sector includes at least 50% of female-founded start-ups within a 4-digit French SIC sector. The mean and number of observations by gender group are reported as well as mean differences and t-statistics between the two groups. The table presents founders' biographical characteristics (panel A), motivation and optimism items (panel B), family structure and composition of the founding team (panel C), start-ups' financing sources (panel D). Variable definitions and data sources are provided in Appendix C.

	Female start-up founders						Male start-up founders					
	Male sector		Female sector		difference	t-stat	Male sector		Female sector		difference	t-stat
	N	Mean	N	Mean			N	Mean	N	Mean		
Panel A. Founder biographical characteristics												
Bachelor's	17840	0.17	6567	0.17	0.01	(1.23)	64314	0.13	3725	0.14	-0.01*	(-1.91)
Master's/PhD	17840	0.27	6567	0.22	0.05***	(8.14)	64314	0.21	3725	0.31	-0.10***	(-13.32)
Elite engineering school	17840	0.03	6567	0.02	0.02***	(7.43)	64314	0.05	3725	0.05	-0.00	(-0.05)
Expert	17840	0.51	6567	0.58	-0.07***	(-9.28)	64314	0.66	3725	0.59	0.07***	(8.38)
Serial	17840	0.23	6567	0.20	0.03***	(4.87)	64314	0.32	3725	0.32	0.01	(0.76)
Panel B. Founder initial motivations and optimism												
High-growth oriented	15664	0.27	6030	0.20	0.07***	(11.65)	55560	0.31	3208	0.29	0.02*	(1.86)
Independence	17840	0.57	6567	0.65	-0.08***	(-11.42)	64312	0.62	3725	0.58	0.04***	(4.23)
Taste	17840	0.42	6566	0.43	-0.01	(-1.18)	64311	0.46	3724	0.44	0.02**	(2.12)
Add earnings	15815	0.22	5696	0.21	0.01*	(1.93)	55065	0.25	3253	0.24	0.01	(1.03)
Opportunity	17840	0.22	6566	0.20	0.02***	(3.13)	64309	0.20	3724	0.21	-0.01	(-1.10)
New idea	17840	0.17	6566	0.15	0.02***	(4.33)	64309	0.16	3724	0.18	-0.02***	(-3.11)
Optimism at start	10838	0.26	4426	0.13	0.13***	(19.18)	37875	0.33	2265	0.25	0.08***	(8.71)
Panel C. Team composition												
Alone	17160	0.69	6391	0.85	-0.15***	(-26.95)	62943	0.73	3572	0.71	0.02**	(2.44)
Spouse	17160	0.15	6390	0.06	0.09***	(23.58)	62941	0.10	3572	0.10	-0.01	(-1.00)
Relatives	17160	0.05	6390	0.02	0.03***	(11.57)	62940	0.05	3572	0.04	0.00	(1.33)
Associates	17160	0.12	6390	0.08	0.04***	(9.57)	62941	0.13	3573	0.15	-0.02***	(-2.69)
Panel D. Start-up financing sources												
External financing	17840	0.50	6567	0.55	-0.05***	(-6.68)	64314	0.51	3725	0.50	0.01	(1.29)
External equity	17840	0.02	6567	0.02	0.01***	(3.52)	64314	0.03	3725	0.03	0.00	(1.18)
Venture capital	15220	0.00	5276	0.00	-0.00	(-0.28)	53192	0.01	3038	0.01	0.00	(1.49)
Bank loans	17840	0.30	6567	0.31	-0.02**	(-2.37)	64314	0.31	3725	0.31	0.00	(0.17)
Personal loans	17840	0.13	6567	0.16	-0.03***	(-5.17)	64314	0.12	3725	0.13	-0.01**	(-2.36)
Non-bank loans	17840	0.11	6567	0.11	-0.00	(-1.00)	64314	0.09	3725	0.08	0.01***	(2.78)
Subsidies	17840	0.16	6567	0.22	-0.06***	(-10.01)	64314	0.17	3725	0.15	0.03***	(4.50)

Table 1.A8. What Makes a Female Entrepreneur?

*Source:* SINE surveys. *Sample:* Repeated cross-section of new firms founded in 2002, 2006, 2010, and 2014. This table predicts the likelihood that a start-up is run by a woman as opposed to a man. Column (1) tests the effect of human capital and includes educational and experience dummy variables. Column (2) adds motivation dummy variables that stem from the question "What are your three main motivations?". Columns (3) and (4) test the same relationship on the subsample of equity-backed startups. Variable definitions are provided in Appendix B. All regressions include (4-digit) French SIC, sector  $\times$  cohort, and zip code fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

	1{Female entrepreneur}			
	Population		Backed with equity	
	(1)	(2)	(3)	(4)
Age $\geq$ 40	-0.001 (-0.10)	-0.005 (-0.63)	-0.013 (-0.78)	-0.012 (-0.72)
French	0.019** (2.35)	0.021*** (2.71)	0.021 (0.87)	0.023 (0.92)
Bachelor's	0.050*** (6.02)	0.052*** (6.25)	0.062** (2.54)	0.063** (2.56)
Master's/PhD	0.022** (2.15)	0.023** (2.33)	-0.011 (-0.51)	-0.013 (-0.58)
Expert	-0.076*** (-9.45)	-0.075*** (-9.54)	-0.062*** (-2.64)	-0.062*** (-2.69)
Serial	-0.082*** (-11.87)	-0.083*** (-12.43)	-0.076*** (-4.47)	-0.077*** (-4.45)
Independence		-0.016*** (-3.36)		0.002 (0.13)
Taste		-0.027*** (-8.71)		-0.016 (-1.14)
Opportunity		0.013*** (3.15)		0.019 (0.91)
New idea		-0.003 (-0.81)		0.019 (1.01)
Sector $\times$ cohort FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
$R^2$	0.184	0.185	0.310	0.311
N	92446	92439	2503	2502

**Table 1.A9. Robustness Test: Conditioning on Getting Other Findings**

*Source:* SINE surveys and tax files. *Sample:* Repeated cross-section of new firms founded in 2002, 2006, 2010, and 2014. The table reports linear probability model estimates and analyzes the effect of gender on the use of external equity for different subsamples of highly motivated entrepreneurs or entrepreneurs who sought other sources of external financing. The dependent variable is a dummy variable that is equal to 1 if the start-up receives VC or other external equity financing. The main independent variable is a dummy variable *Female*. The relationship is estimated on different subsamples: the subsample of *High-growth oriented* entrepreneurs (column 1), entrepreneurs who raised bank debt for their start-up (column 2), entrepreneurs who have a personal loan (column 3) and entrepreneurs who received a grant/subsidy (column 4). They also include human capital and start-up controls, that are defined in the variable appendix. All models include county and 4-digit French SIC sector  $\times$  cohort-year fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Sub-samples:	Is High-growth oriented (1)	Has a Bank loan (2)	Has a Personal loan (3)	Has a Subsidy (4)
Female	-0.0097*** (-4.50)	-0.0068*** (-2.66)	-0.0123*** (-3.35)	-0.0088*** (-3.52)
Age $\geq$ 40	0.0043* (1.80)	0.0046** (2.14)	0.0069** (2.24)	0.0071*** (2.96)
French	-0.0032 (-0.94)	-0.0045 (-1.00)	-0.0026 (-0.52)	-0.0064 (-1.41)
Bachelor's	-0.0019 (-0.69)	0.0039 (1.19)	0.0053 (1.28)	0.0000 (0.01)
Master's/PhD	0.0163*** (6.71)	0.0191*** (6.19)	0.0115** (2.59)	0.0183*** (4.36)
Expert	0.0020 (1.05)	0.0013 (0.45)	0.0018 (0.60)	-0.0041 (-1.28)
Serial	0.0207*** (8.66)	0.0190*** (7.81)	0.0075** (2.23)	0.0059* (1.83)
Incorporated	0.0184*** (6.97)	0.0147*** (5.87)	0.0291*** (8.57)	0.0218*** (7.33)
Tangible/total assets	-0.0091** (-1.99)	-0.0026 (-0.52)	-0.0053 (-1.06)	-0.0012 (-0.35)
Ln(total assets)	0.0049*** (6.56)	0.0044*** (4.21)	0.0018* (1.72)	0.0056*** (3.90)
Sector $\times$ cohort FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
$R^2$	0.073	0.065	0.091	0.097
N	35220	28452	11394	15809

**Table 1.A10. Robustness Test: Controlling for Invested Capital at Creation**

*Source:* SINE survey and tax files. *Sample:* New firms founded in 2002, 2006, 2010, and 2014. The table reports linear probability model estimates and analyzes the effect of gender the use of external equity, controlling for entrepreneur's invested capital at start. Panel A reports the distribution of start-ups by gender and amount of capital invested at creation. The dependent variable of all models in panel B are a dummy variable that is equal to 1 if the start-up receives VC or other external equity financing. The main independent variable is a dummy variable *Female*. All models in panel B include human capital and start-up controls, that are defined in the variable appendix. They also include county and 4-digit French SIC sector  $\times$  cohort-year fixed effects interacted with the invested capital buckets defined in panel A. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Panel A. Distribution of starting capital

Starting capital	Male	Female	All
<2000	12,667	5,881	18,548
%	18.60	24.09	20.05
[2,000;4,000[	7,886	2,852	10,738
%	11.58	11.68	11.61
[4,000;8,000[	11,288	3,463	14,751
%	16.57	14.19	15.94
[8,000;16,000[	12,428	3,902	16,330
%	18.25	15.99	17.65
[16,000;40,000[	11,903	4,115	16,018
%	17.48	16.86	17.31
[40,000;80,000[	5,185	2,022	7,207
%	7.61	8.28	7.79
$\geq 80,000$	6,755	2,175	8,930
%	9.92	8.91	9.65

Panel B. Regressions with starting capital buckets

	External equity (1)	Venture capital (2)	Bank loans (3)	Subsidies (4)
Female	-0.0064*** (-3.81)	0.0052 (0.91)	0.0128*** (3.56)	0.0064 (1.22)
Human capital controls	Yes	Yes	Yes	Yes
Start-up controls	Yes	Yes	Yes	Yes
Sector $\times$ cohort $\times$ investment bucket FE	Yes	Yes	Yes	Yes
$R^2$	0.353	0.364	0.516	0.491
N	49230	49230	49230	49230

**Table 1.A11. Capital Structure at Start**

*Source:* SINE survey and tax files. *Sample:* Repeated cross-section of new firms founded in 2002, 2006 or 2010 depending on the dependent variable. The table reports linear probability model estimates and analyzes the effect of gender and other entrepreneurs' characteristics on the % of personal resources (columns 1 and 2) in the capital structure, the % of debt (columns 3 and 4) and the % of other external resources (columns 5 and 6). All models include the *Female* dummy, human capital and start-up controls, that are defined in the variable appendix. In odd columns, regressions are estimated on the full population of entrepreneurs. In even columns, regressions are estimated on the subsample of start-up founders who received external equity. All models include county and 4-digit French SIC sector  $\times$  cohort-year fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Dependent variable:	% personal resources		% debt		% other resources	
Sample:	All	Backed	All	Backed	All	Backed
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-1.1720*	0.4051	1.0433*	-1.4738	-0.1309	-0.2055
	(-1.96)	(0.16)	(1.82)	(-0.50)	(-0.47)	(-0.08)
Age $\geq$ 40	2.6402***	-2.6534	-3.1222***	-2.3591	0.4288*	3.3951*
	(4.61)	(-1.19)	(-5.69)	(-1.33)	(1.93)	(1.75)
French	-9.2285***	-9.2582**	8.1941***	5.9457*	0.6045	-2.2421
	(-11.65)	(-2.15)	(13.13)	(1.75)	(1.34)	(-0.37)
Bachelor's	0.7344	4.6637	0.9474	3.1749	-0.0949	-8.2480**
	(1.01)	(1.45)	(1.60)	(1.16)	(-0.31)	(-2.36)
Master's/PhD	3.6568***	0.6361	-2.5524***	-6.3968***	0.3874	6.8110**
	(5.27)	(0.26)	(-4.80)	(-3.45)	(1.46)	(2.55)
Expert	-1.3598***	-1.3392	1.2416**	-0.1471	-0.2486	0.9126
	(-2.77)	(-0.64)	(2.47)	(-0.08)	(-1.09)	(0.36)
Serial	4.0369***	3.1044	-1.7107***	-0.0377	-1.8789***	-1.0729
	(7.81)	(1.29)	(-3.67)	(-0.02)	(-7.95)	(-0.44)
Incorporated	1.2440*	-2.4067	2.5089***	3.0697	-1.7190***	1.8213
	(1.70)	(-0.76)	(4.12)	(1.07)	(-6.42)	(0.57)
Sector $\times$ cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
County	Yes	Yes	Yes	Yes	Yes	Yes
$R^2$	0.092	0.228	0.151	0.303	0.026	0.219
N	46292	1305	40567	1136	33680	996
Available cohorts	2002 and 2010		2006 and 2010		2010	

**Table 1.A12. Robustness Test: Conditioning on Getting Other  
Fundings**

*Source:* SINE survey and tax files. *Sample:* New firms founded in 2002, 2006, 2010, and 2014. The table reports linear probability model estimates and analyzes the effects of the interaction between gender and having obtained other financing sources on the use of external equity. The dependent variable is a dummy variable that is equal to 1 if the start-up receives VC or other external equity financing. The main independent variable is a dummy variable *Female*. It is interacted with other external financing sources: bank loan (column 1), personal loan (column 2), non-bank loan (column 3) and subsidy (column 4). All models include human capital and start-up controls, that are defined in the variable appendix. They also include county and 4-digit French SIC sector  $\times$  cohort-year fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Dependent variable	External equity			
	Bank loans (1)	Personal loans (2)	Non-bank loans (3)	Subsidies (4)
Female	-0.0045*** (-3.92)	-0.0046*** (-3.85)	-0.0055*** (-4.74)	-0.0052*** (-4.12)
Other external funding	0.0089*** (6.15)	-0.0005 (-0.26)	0.0046** (2.09)	-0.0028 (-1.25)
Female x Other external funding	-0.0032 (-1.37)	-0.0061** (-2.11)	-0.0004 (-0.10)	-0.0016 (-0.61)
Human capital controls	Yes	Yes	Yes	Yes
Start-up controls	Yes	Yes	Yes	Yes
Sector $\times$ cohort FE	Yes	Yes	Yes	Yes
County	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.044	0.043	0.043	0.043
N	92336	92336	92336	92336

**Table 1.A13. The Effects of Gender Stereotypes - Models without Fixed Effects**

*Source:* SINE surveys and tax files. *Sample:* Repeated cross-section of new firms founded in 2002, 2006, 2010, and 2014. The table reports OLS estimates, replicates table 4 without fixed effects and displays . The dependent variables is a dummy variable that is equal to 1 if the start-up receives VC or other external equity financing. *Female-dominated sector*, Entrepreneurs is a dummy variable that takes a value of 1 if more than 50% of new start-ups within a 4-digit French SIC sector are female-founded (column 1). Other measures of *Female-dominated sector* also capture within sector female representativeness and are based on the % of new female-led businesses (column 2), % of female small business owners (column 3) and % of female CEOs (column 4). Human capital controls include entrepreneur's age and French citizenship, education and experience dummy variables. Start-up controls include an incorporation status dummy variable, the ratio of tangible assets and the logarithm of firm's total assets. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at the 10, 5, and 1% levels, respectively.

Dependent variables:	Venture Capital			
	Entrepreneurs (1)	New business owners (2)	All business owners (3)	CEOs (4)
Measures of female-dominated sector:				
Female	-0.0074*** (-6.23)	-0.0077*** (-6.45)	-0.0087*** (-7.39)	-0.0084*** (-6.27)
Female-dominated sector	-0.0037* (-1.83)	-0.0025 (-1.00)	-0.0036* (-1.90)	-0.0016 (-0.88)
F × F-dominated sector	0.0082*** (3.24)	0.0064** (2.18)	0.0096*** (3.69)	0.0054** (2.29)
Human capital controls	Yes	Yes	Yes	Yes
Start-up controls	Yes	Yes	Yes	Yes
Sector controls	No	No	No	No
Sector FE	No	No	No	No
Cohort FE	No	No	No	No
county FE	No	No	No	No
$p - val:$	= 0	0.0185	0.0772	0.0057
$R^2$	0.011	0.011	0.011	0.011
N	92574	92480	92387	92574

**Table 1.A14. The Effects of Gender Stereotypes on Venture Capital Only**

*Source:* SINE surveys and tax files. *Sample:* Repeated cross-section of new firms founded in 2002, 2006, 2010, and 2014. The table reports linear probability model estimates and analyzes the effect of gender stereotypes on the use of different financing sources. The dependent variable is a dummy variable that is equal to 1 if the start-up receives VC financing. *Female-dominated sector* (Entrepreneurs) in column 1 is a dummy variable that takes a value of 1 if more than 50% of new start-ups within a 4-digit French SIC sector are female-founded. Other measures of *Female-dominated sector* also capture within sector female representativeness and are based on the % of new female-led businesses (column 2), the % of female small business owners (column 3) and the % of female CEOs (column 4). Control variables at the 4-digit French SIC level are the within-sector percentage of female-founded start-ups, Herfindahl index and the logarithm of total sector sales. Human capital controls include entrepreneur's age and French citizenship, education and experience dummy variables. Start-up controls include an incorporation status dummy variable, the ratio of tangible assets and the logarithm of firm's total assets. All models include county, 4-digit French SIC sector and cohort-year fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at the 10, 5, and 1% levels, respectively.

Dependent variables:	Venture Capital			
	Entrepreneurs	New business owners	All business owners	CEOs
Measures of female-dominated sector:	(1)	(2)	(3)	(4)
Female	-0.0027*** (-4.40)	-0.0033*** (-4.41)	-0.0033*** (-5.04)	-0.0033*** (-5.19)
Female-dominated sector	-0.0007 (-0.24)	-0.0058 (-1.20)	0.0027 (0.79)	0.0007 (0.26)
F × F-dominated sector	0.0031* (1.86)	0.0024* (1.94)	0.0041*** (2.63)	0.0037** (2.43)
Additional sector controls	Yes	Yes	Yes	Yes
Human capital controls	Yes	Yes	Yes	Yes
Start-up controls	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
$R^2$	0.022	0.021	0.022	0.022
N	76859	76716	76769	76859

**Table 1.A15. Future Performance of Non-backed Start-ups**

*Source:* SINE surveys, tax files and employers' payrolls. *Sample:* Panel (columns 1-4) and repeated cross sections (columns 5-8) of *non* equity-funded start-ups founded in 2002, 2006, 2010, and 2014. The table reports OLS estimates and examines how funded minority entrepreneurs perform relative to entrepreneurs from the dominant gender group within and across sectors. Panel A examines performance of female entrepreneurs in male dominated sectors. Panel B examines performance of male entrepreneurs in female-dominated sectors. The *Female* gender is interacted with *Male-dominated sector* which is a dummy variable that takes a value of 1 if more than 50% of firms within a 4-digit French SIC are male-founded. The dependent variables are measures of future corporate development, performance and risk, and include the *Number of employees* (column 1), the ratio of *Sales over assets* (column 2), *% foreign sales* (column 3), the *ROA* (column 4), from one year after creation up to five years onwards. Other measures of performance include dummy variables *Survival 5 years* and *Survival 3 years* that takes the value one if the firm has survived at least 5 or 3 years, respectively (columns 5-6). Measures of corporate risk include firms' volatility of employment and sales, respectively, within firm from 1 year after creation up to 5 years (columns 7 and 8). All regressions include human capital and start-up fixed effects, as well as county, 4-digit French SIC sector and cohort-year fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Panel A. Female entrepreneurs in male-dominated sectors								
Dependent variables:	N. Employees (1)	Sales over assets (2)	% foreign sales (3)	ROA (4)	Survival 5 years (5)	Survival 3 years (6)	Vol. em- ployment (7)	Vol. sales (8)
Female	-0.3396*** (-13.27)	0.3696*** (3.95)	0.0090*** (4.37)	0.0706*** (9.50)	-0.0067 (-0.68)	0.0004 (0.04)	-9.9415* (-1.78)	-9.9415* (-1.78)
Male-dominated sector	-0.1236*** (-3.19)	0.0651 (0.77)	-0.0262*** (-8.48)	0.0540*** (6.00)	0.0292** (2.14)	0.0287** (2.28)	7.5571 (0.52)	7.5571 (0.52)
Female × M-dominated sector	0.3977*** (14.09)	-0.3252*** (-3.34)	-0.0106*** (-4.81)	-0.1030*** (-12.83)	-0.0285*** (-2.64)	-0.0333*** (-3.34)	-0.5519 (-0.03)	-0.5519 (-0.03)
Human capital controls	Yes	Yes	Yes	Yes	Yes	Yes		
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.284	0.069	0.153	0.081	0.059	0.044	0.080	0.080
N	240301	266794	266803	267606	89701	89701	76082	76082
Panel B. Male entrepreneurs in female-dominated sectors								
Dependent variables:	N. Employees (1)	Sales over assets (2)	% foreign sales (3)	ROA (4)	Survival 5 years (5)	Survival 3 years (6)	(7)	(8)
Male	-0.0512 (-1.40)	-0.0445 (-1.58)	0.0017** (1.98)	0.0268*** (5.98)	0.0352*** (8.23)	0.0329*** (8.24)	10.4934 (0.71)	10.4934 (0.71)
Female-dominated sector	-0.1002 (-0.87)	0.2601*** (3.23)	0.0368*** (12.19)	0.0723*** (5.48)	-0.0007 (-0.05)	0.0045 (0.36)	-7.0051 (-0.44)	-7.0051 (-0.44)
M × F-dominated sector	0.1853** (2.36)	-0.3252*** (-3.34)	-0.0106*** (-4.81)	-0.1527*** (-10.47)	-0.0285*** (-2.64)	-0.0333*** (-3.34)	-0.5519 (-0.03)	-0.5519 (-0.03)
Human capital controls	Yes	Yes	Yes	Yes	Yes	Yes		
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.254	0.069	0.153	0.089	0.059	0.044	0.080	0.080
N	242346	266794	266803	272280	89701	89701	76082	76082

**Table 1.A16. Future Performance of Bank-financed Start-ups**

*Source:* SINE surveys, tax files and employers' payrolls. *Sample:* Panel (columns 1-4) and repeated cross sections (columns 5-8) of bank debt-financed start-ups founded in 2002, 2006, 2010, and 2014. The table reports OLS estimates and examines how funded minority entrepreneurs perform relative to entrepreneurs from the dominant gender group within and across sectors. Panel A examines performance of female entrepreneurs in male dominated sectors. Panel B examines performance of male entrepreneurs in female-dominated sectors. The *Female* gender is interacted with *Male-dominated sector* which is a dummy variable that takes a value of 1 if more than 50% of firms within a 4-digit French SIC are male-founded. The dependent variables are measures of future corporate development, performance and risk, and include the *Number of employees* (column 1), the ratio of *Sales over assets* (column 2), *% foreign sales* (column 3), the *ROA* (column 4), from one year after creation up to five years onwards. Other measures of performance include dummy variables *Survival 5 years* and *Survival 3 years* that takes the value one if the firm has survived at least 5 or 3 years, respectively (columns 5-6). Measures of corporate risk include firms' volatility of employment and sales, respectively, within firm from 1 year after creation up to 5 years (columns 7 and 8). All regressions include human capital and start-up fixed effects, as well as county, 4-digit French SIC sector and cohort-year fixed effects. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

Panel A. Female entrepreneurs in male-dominated sectors						
Dependent variables:	N. Employees	Sales over assets	% foreign sales	ROA	Survival 5 years	Survival 3 years
Vol. employment	Vol. sales					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.2355*** (-2.86)	0.4679*** (4.74)	-0.0009 (-0.39)	0.1529*** (9.03)	-0.0088 (-0.52)	0.0023 (0.16)
Male-dominated sector	-0.0921 (-0.48)	-0.1674 (-1.48)	-0.0193*** (-4.61)	0.0675*** (3.51)	-0.0062 (-0.27)	-0.0042 (-0.20)
F × M-dominated sector	0.3024*** (2.72)	-0.3392*** (-3.19)	-0.0000 (-0.00)	-0.1740*** (-9.69)	-0.0219 (-1.19)	-0.0233 (-1.43)
Human capital controls	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.385	0.080	0.076	0.090	0.092	0.070
N	90248	95430	95460	97501	28679	28679
Panel B. Male entrepreneurs in female-dominated sectors						
Dependent variables:	N. Employees	Sales over assets	% foreign sales	ROA	Survival 5 years	Survival 3 years
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.0669 (-0.91)	-0.1287*** (-3.19)	0.0010 (0.88)	0.0283*** (5.86)	0.0307*** (4.07)	0.0210*** (3.07)
Female-dominated sector	-0.2103 (-1.18)	0.5066*** (4.61)	0.0193*** (5.02)	0.0765*** (5.03)	0.0281 (1.23)	0.0275 (1.36)
M × F-dominated sector	0.3024*** (2.72)	-0.3392*** (-3.19)	-0.0000 (-0.00)	-0.1134*** (-9.53)	-0.0219 (-1.19)	-0.0233 (-1.43)
Human capital controls	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.385	0.080	0.076	0.094	0.092	0.070
N	90248	95430	95460	96560	28679	28679

**Table 1.A17. Descriptive Statistics of the VentureXpert Deals**

*Source:* VentureXpert, tax files and employer payrolls. *Sample:* Equity backed firms available in VentureXpert from 2002 to 2014 that could be linked to the tax files and employer payrolls databases. The table reports the distribution of characteristics of companies that receive PE, VC, CVC, and angel investment financing (Panel A), corporate outcomes of those companies (Panel B), and characteristics of the PE and VC investment firm (Panel C).

	N	Mean	Sd	P25	P50	P75
Panel A. Entrepreneurs's characteristics						
Female entrepreneur	1691	0.149	0.357	0.000	0.000	0.000
Age	1623	35.77	275	34.92	0.85*	(1.94)
Panel B. Deal's characteristics						
<u>Type:</u>						
Venture capital	1691	0.44	297	0.38	0.06**	(2.07)
Private equity	1691	0.59	297	0.61	-0.01	(-0.40)
Angel	1691	0.15	297	0.18	-0.03	(-1.22)
CVC	1691	0.06	297	0.03	0.03**	(2.26)
Early stage	1691	0.30	297	0.23	0.07**	(2.42)
Total amount invested	1189	189.27	194	138.71	50.55	(0.86)
Total amount invested in 1st round	1151	120.46	193	97.90	22.56	(0.47)
<u>Exit:</u>						
IPO	1331	0.06	221	0.05	0.01	(0.44)
M&A	1609	0.22	285	0.27	-0.04	(-1.52)
Bankruptcy	1691	0.01	297	0.01	-0.00	(-0.49)
<u>Growth and performance: N. employees</u>						
Sales over assets	1538	0.88	264	0.94	-0.06	(-1.00)
ROA	1538	-0.13	264	-0.16	0.03	(1.22)
Tangible/total assets	1538	0.08	264	0.07	0.01	(0.80)
Ln(total assets)	1538	8.07	264	7.58	0.49***	(4.39)
Panel C. PE and VC investment firms characteristics						
Female VC	1240	0.08	225	0.09	-0.01	(-0.41)
Age VC	1240	37.85	225	38.22	-0.37	(-0.73)
VC has one women among top 3 employees	1691	0.66	297	0.67	-0.00	(-0.13)
N. employees	1240	25.56	225	20.10	5.46**	(2.35)

**Table 1.A18. Gender Stereotypes and Types of External Equity Investors**

*Source:* VentureXpert, tax files and employer payrolls databases. *Sample:* Equity backed firms available in VentureXpert from 2002 to 2016 that could be linked to the tax files and employer payrolls databases. The table reports linear probability model estimates and replicates out of sample the main results displayed in table 4. The dependent variables are as follows: *External equity* (column 1) is a dummy variable that is equal to 1 if the firm receives any type of external equity financing. *Venture capital* (column 2) is equal to 1 if the firm receives venture capital financing. *Private equity* (column 3) is equal to 1 if the firm receives any other type of private equity financing. *Angel* (column 4) is equal to 1 if the firm receives angel investment financing. *CVC* (column 5) is equal to 1 if the firm receives corporate venture capital financing. Female CEOs are identified using an occupation code, or if not available, as the highest paid employee in employer payrolls database. *Female-dominated sector* is a dummy variable that takes the value of 1 if more than 50% of firms within a 4-digit French SIC sector are female-led. Control variables include the ratio of tangible assets, the logarithm of firm's total assets, the sector Herfindhal index and sector size. All models sector fixed effects and cohort fixed effects defined in this case by the intersection between year and age bucket. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

	External equity (1)	Venture capital (2)	Private equity (3)	Angel (4)	CVC (5)
Female	-0.00031*** (-3.43)	-0.00020*** (-3.26)	-0.00012** (-2.08)	-0.00007* (-1.76)	-0.00003** (-2.01)
Female-dominated sector	-0.00051 (-1.38)	-0.00142** (-2.38)	0.00028 (0.75)	0.00021 (0.30)	-0.00044 (-0.90)
Female × F-dominated sector	0.00028** (2.28)	0.00021*** (2.80)	0.00007 (0.87)	0.00010** (2.04)	0.00001 (0.37)
Tangible/total assets	-0.00025*** (-3.03)	-0.00013** (-2.57)	-0.00012** (-2.54)	-0.00007** (-2.58)	0.00001 (0.48)
Ln(total assets)	0.00058*** (5.24)	0.00024*** (4.44)	0.00038*** (5.71)	0.00009*** (3.44)	0.00003*** (3.34)
Sector controls	Yes	Yes	Yes	Yes	Yes
Cohort (Year times Age bucket) FE	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.008	0.010	0.006	0.005	0.004
N	1332717	1332717	1332717	1332717	1332717

**Table 1.A19: Entrepreneur Gender and Alternative Measures of Performance**

The table reports linear probability model estimates and tests the effect of stereotypes on the likelihood of different types of exits. The dependent variables are *IPO* (column 1) that is equal to 1 if the firm exits by IPO in the following years; *M&A* (column 2) which is equal to 1 if the firm is acquired in the following years; and *Bankruptcy* (column 3) which is a dummy variable that takes the value 1 if the firm files for bankruptcy in the following year. Note that only firms that received external equity investment are included in the sample. Robust standard errors are reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at the 10, 5, and 1% levels, respectively. *Source*: VentureXpert. PE- and VC-backed deals available in VentureXpert from 2002 to 2016 that could be linked to the tax files and employer payrolls databases.

Panel A. Female entrepreneurs in male-dominated sectors			
	IPO (1)	M&A (2)	Bankruptcy (3)
Constant	0.084*** (3.90)	0.277*** (9.40)	0.009 (1.42)
Female	-0.066** (-2.40)	-0.017 (-0.30)	-0.009 (-1.42)
Male-dominated sector	-0.019 (-0.82)	-0.040 (-1.26)	-0.001 (-0.12)
F × Male-dominated sector	0.079** (2.19)	-0.053 (-0.81)	0.012 (1.15)
$R^2$	0.003	0.004	0.001
N	1256	1643	1643

Panel B. Male entrepreneurs in female-dominated sectors			
	IPO (1)	M&A (2)	Bankruptcy (3)
Constant	0.077*** (3.60)	0.167*** (6.09)	0.011 (1.42)
Male	-0.012 (-0.54)	0.070** (2.33)	-0.003 (-0.36)
Female-dominated sector	-0.060** (-2.16)	0.093 (1.63)	-0.011 (-1.42)
Male × F-dominated sector	0.079** (2.19)	-0.053 (-0.81)	0.012 (1.15)
$R^2$	0.003	0.004	0.001
N	1256	1643	1643

**Table 1.A20: Entrepreneur Gender, Amount Invested and Rounds of Financing**

*Source:* VentureXpert. *Sample:* PE- and VC-backed deals available in VentureXpert from 2002 to 2016 that could be linked to the tax files and employer payrolls databases. The table reports OLS estimates and examines the effect of gender stereotypes on amount invested by equity investors. The dependent variables are the *Logarithm (total amount received)* and *Number of rounds*. The variable *Female* entrepreneur is interacted with the dummy variable *Female-dominated sector*, that takes the value 1 if more than 50% of start-ups in the 4-digit SIC sector are female-led. The models include sector fixed effects and cohort fixed effects defined as the intersection between the year and the age bucket. Clustered standard errors at the sector level are reported in parentheses. \*, \*\*, and \*\*\* indicate significantly different from zero at the 10, 5, and 1% levels, respectively.

	Ln(Amount)	Number of rounds
	(1)	(2)
Female	-0.130 (-0.84)	-0.029 (-0.49)
Female-dominated sector	-0.518*** (-2.82)	-0.244*** (-2.66)
Female × F-dominated sector	0.129 (0.31)	0.128 (1.15)
Cohort FE	Yes	Yes
Sector FE	Yes	Yes
$R^2$	0.078	0.098
N	1339	1929

### 1.11.2. Scrapping

In this appendix, I describe the procedure used to find one-to-one correspondences between firms involved in deals reported in the commercial database Thomson VentureXpert and the French administrative data.<sup>37</sup> Databases maintained by the French Bureau of Statistics (INSEE) contain firm standardized 9-digit identifiers, called SIREN. These firm standardized identifiers are not provided in commercial databases. Commercial databases usually include the firms' name, address, zip code, and country. I use this information as the input for a Python web-crawler to look for a firm's name and address on two websites: (i) [www.bodacc.fr](http://www.bodacc.fr) (*Bulletin Officiel des Annonces Civiles et Commerciales*), a governmental website that has collected and published official notifications involving French companies since 2008, and (ii) [www.societe.com](http://www.societe.com), a commercial website that aggregates and reshapes information about French companies from various sources (mostly from INSEE and Bodacc.fr). Both websites are supposed to cover the universe of French firms. The web-crawler is built using the Python packages <https://www.crummy.com/software/BeautifulSoup/BeautifulSoup> and <https://www.seleniumhq.org/>. The output is a list with names and corresponding information. The next step consists of checking the quality of the matches. First, I drop observations with several matches that do not report an address, city, zip code or any other information. Second, I impose a maximum Jaró-Winkler string distance between the original and retrieved names of 0.8 for both the name and address and drop the matches that do not meet this restriction.<sup>38</sup> Third, in the case where there are still several matches, I keep the correspondence with the highest Jaró-Winkler string distance. Results of the matching procedure by are given in table 1.A21. Using a sample of deals from 2000 to 2018 retrieved from the database VentureXpert, the scrapping procedure succeeded in finding a one-to-one correspondence for 76% of the targets involved in the sample of VC-PE deals and for 80% of firms targeted by angel investors.

---

<sup>37</sup>A similar procedure is used for a sample of M&A deals retrieved from Thomson Reuters SDC in Beaumont, P., Hebert, C., and Lyonnet, V., 2018, "Build or buy? Human capital and corporate diversification", Working Paper.

<sup>38</sup>This distance measures the number of characters in common between strings with the idea that differences near the start of the string are more significant than differences near the end of the string.

**Table 1.A21. Scrapping Success Rate by Deal Year**

*Source:* VentureXpert *Sample:* Private equity, venture capital and angel investment deals downloaded from VentureXpert from January 2000 to December 2015. The table reports the number of target firms involved in a private equity deals downloaded from VentureXpert and the number of target firms for which the web-scraping procedure has succeeded to find a unique SIREN identifier.

Year	PE and VC		Angel	
	Downloaded	Final	Downloaded	Final
2002	191	124	21	15
2003	456	308	27	17
2004	482	309	55	36
2005	283	197	59	38
2006	305	212	61	43
2007	340	244	99	77
2008	339	264	68	53
2009	275	222	60	48
2010	372	316	75	65
2011	386	316	119	100
2012	368	306	129	101
2013	315	251	82	71
2014	309	264	110	94
2015	392	328	158	141
Total	4,813	3,661	1,123	899
Success rate		0.761		0.800

## Chapter 2

# Build or Buy? Human Capital and Corporate Diversification

\*

with Paul Beaumont (Université Paris Dauphine) and Victor Lyonnet (Ohio State University)

---

\*This chapter benefited from the comments and discussions with Manuel Adelino, Philippe Aghion, Fabio Braggion, Edouard Challe, Espen Eckbo, Simon Gervais, Edith Ginglinger, John Graham, Denis Gromb, Johan Hombert, Julien Martin, Adrien Matray, Isabelle Méjean, Mario Milone, Tamara Nefedova, Paige Ouimet, Gordon Phillips, Manju Puri, Amiyatosh Purnanandam, Adriano Rampini, Luc Renneboog, David Robinson, René Stulz, Geoffrey Tate, Vish Viswanathan, Ramona Westermann, Sheng-Jun Xu, as well as seminar participants at CREST, Duke University, the 5th ECGC workshop on Governance and Control, HEC Paris, INSEE, the 28th Mitsui Labor and Finance Symposium, NFA, Ohio State University, Université Paris-Dauphine, UT Austin Corporate Finance Forum, Tilburg University. This chapter was awarded best student paper award at the 5th ECGC workshop on Governance and Control.

## **Abstract**

Why do some firms enter a new sector by acquiring an existing company ("buy"), while others do so using their existing resources ("build")? Using a novel data set constructed by merging French employer payrolls with commercial M&A data sets, we show that firms are more likely to buy when their existing workforce does not include skills needed in the sector of entry. This relationship is more pronounced when labor market frictions make it difficult to hire key workers. Firms that enter by building realize lower entry sales when their existing workforce is not adapted to the sector of entry, especially in the presence of labor market frictions. Our results suggest that firms buy to acquire their targets' human capital when adapting their existing workforce is too costly.

## 2.1. Introduction

More than twelve thousand cross-industry merger and acquisition (M&A) deals occurred worldwide in 2017, and amounted to a total value of more than 900 billion dollars.<sup>2</sup> Cross-industry M&As allow firms to diversify, obtain specific assets and access new markets.<sup>3</sup> However, the magnitude of these figures can be misleading, concealing the fact that most diversifications result from firms entering new sectors directly by building on their existing capabilities. Why do some firms enter a new sector by acquiring an existing company (“buy”), while others do so using their existing resources (“build”)? This question has implications for understanding the role played by M&As in the reallocation of resources in the economy.

In this paper, we compare firms that enter a new sector by building on existing resources to firms that buy an existing company operating in the sector of entry. We focus on the role of human capital as a key resource to successfully operate in the new sector. Although our analysis applies to other assets, the availability of employee-level information and the existence of specific labor market frictions motivate our focus on human capital. We study the costs and benefits associated with each alternative. When a firm buys to enter a new sector, it has to incur both the costs of acquiring and restructuring the target, but the acquiring company also secures access to the target’s productive resources. When a firm builds on its existing resources to enter a new sector, it faces the costs of hiring new workers to complement its existing workforce.<sup>4</sup> Therefore, the more key workers there are in a firm’s existing workforce, the lower the adjustment costs to the new sector, and the more profitable it is for a firm to build.

We find that firms that have a workforce more adapted to the sector of entry are more likely to “build” than to “buy”. The relation between internal human capital and the mode of entry in a new sector is stronger when specific skills are difficult to obtain on the external job market, i.e., when key worker occupations are in short supply. Firms that build have lower entry sales when their workforce is not adapted to the sector of entry, especially when key workers are in short supply. These findings are consistent with the idea that firms must search for key assets to expand into a new sector. When search costs increase, buying an existing firm becomes relatively

---

<sup>2</sup>Cross-sector M&As represent approximately 30% of the universe of M&A deals in both number and volume. See <https://home.kpmg/xx/en/home/insights/2018/04/m-and-a-predictor-2018-annual-report.html>KPMG M&A Predictor 2018 Annual Report.

<sup>3</sup>The motives for horizontal M&As usually differ from those of vertical and conglomerate diversifications; they are often motivated by market power consolidation within a sector (e.g., Eckbo, 1983; Farrell and Shapiro, 1990).

<sup>4</sup>Maksimovic, Phillips and Prabhala (2011) shows that acquiring firms enter a costly restructuring process and sell or close 46% of the plants they buy. Hiring costs have been estimated to account for between one-quarter and one-half of wage payments and to increase in the specificity of skills (Abowd and Kramarz, 2003; Blatter, Muehleman and Schenker, 2012).

more attractive than building from scratch and hiring new workers.<sup>5</sup>

One key challenge in testing this hypothesis is defining human capital and measuring it at the firm level. Human capital is neither directly observable nor easily defined. To overcome this challenge, we propose a measure based on a model of diversification with endogenous choice of teams. In the model, a firm hires workers with different occupations that relate to different sector-specific skills. If the firm chooses to build, it can draw workers from its existing pool of workers (internal labor market) or from the external job market. Instead, if the firm chooses to buy, it can select workers in the workforce of the target firm but has to pay fixed costs for the acquisition of another firm. The model yields two predictions. First, in equilibrium, the firm chooses to build when its existing *internal human capital* is adapted to operate in the sector of entry. The firm instead buys when its workforce does not include the key worker occupations needed to operate in the sector of entry and does so despite restructuring costs. Second, the relationship between human capital and the decision to build or buy is stronger when key workers are in short supply in the external job market.

The model yields a firm-level measure of human capital that stems from the relation between the fraction of the wage bill that goes to a given type of worker and the contribution of such workers to the firm's output. Empirically, we construct our measure of human capital in two steps. First, we use the French matched employer-employee dataset. In the spirit of Abowd, Kramarz and Margolis (1999), we regress the (log) fraction of the wage bills of all firms in a sector that goes to a given occupation on occupation  $\times$  sector fixed effects. We interpret the occupation  $\times$  sector fixed effects estimates as a score reflecting the sector-specific human capital of worker occupations. The higher the fixed effect of a given occupation in a sector is, the larger the *average* share of a firm's wage bill that goes to the corresponding worker type. Second, we aggregate worker human capital at the firm level. A firm's internal human capital is the average of the (exponentiated) occupation  $\times$  sector fixed effects present in the firm's workforce. Our measure of human capital captures the extent to which the existing workforce is adapted to the sector of entry.<sup>6</sup>

Our measure of human capital has several advantages. First, because the French matched employer-employee dataset contains a detailed occupation code at the individual level, our measure of human capital encompasses various determinants of individuals' human capital, including skills, education and experience (Becker, 1957;

---

<sup>5</sup>Cisco's acquisition strategy illustrates this trade-off. In 1997, a Cisco analyst described the strategy of the firm as follows: "in today's economy, building work teams from scratch can be yesterday's luxury. So, when you can't build fast enough, you buy". By 2017, Cisco had undertaken more than 200 M&As to "provide a capability, acceleration potential or earlier sectoral entry compared to partnering or developing in-house" (Wysocki, 1997; Romanski, 2017).

<sup>6</sup>Hereafter, we refer to this variable as "internal human capital" without mentioning that it is specific to the sector of entry.

Gibbons and Waldman, 2004; Autor and Dorn, 2009). Second, the fixed effects strategy allows us to rank occupations within sectors following their average contribution to firms' output. Third, our measure of human capital does not reflect firm-specific unobservables, like personnel policies or the influence of unions that may alter the returns to observable and unobservable dimensions of human capital.

Another empirical challenge for our paper is to identify how firms diversify. We use a detailed breakdown of firms' sales across sectors to identify entries in new sectors.<sup>7</sup> A firm enters a new sector if (i) at least one of its subsidiaries begins selling in the new sector and (ii) none of the other subsidiaries already operates in the sector. We identify buy entries by linking M&A deals retrieved from SDC Platinum and Bureau van Dijk Zephyr to the French administrative data.<sup>8</sup> The entry is a "buy" if the subsidiary that begins selling in the new sector has been acquired by the firm. By contrast, a firm "builds" if entry is made through an existing subsidiary. Our final dataset consists of 75,000 build or buy decisions in France from 2003 to 2014.

We provide the first cross-industry statistics on firms' decisions to build or buy. We find that 98% of entries in a new sector consisted of firms that build on their internal resources. The figure is 90% when weighting by entry sales, meaning that diversification by acquisition represent 10% of the universe of corporate diversification. At the time of entry, firms invest on average one million euros and approximately half of the firms in our sample continue to operate in the sector of entry after one year. The median (mean) entry sales are equal to 270,000 euros (2.8 million euros).

Our main test analyzes whether firms' existing workforce composition explains the decision to build or buy. We compare firms that operate within the same sector of origin and that then diversify within the same sector of entry in the same year. This specification neutralizes potential unobservable time-varying synergies between the sector of entry and that of origin. We find that firms are more likely to buy when their human capital is not adapted to the sector of entry. A one-standard-deviation decrease in human capital is associated with a 1.1 percentage-point increase in the likelihood of buying. This relationship is sizable, equal to 50% of the unconditional probability of buying. This finding holds when adding firm fixed effects, or when controlling for a wide variety of firm characteristics such as size, profitability, capital intensity, cash holdings and other assets.

We then exploit cross-sectional variations in the diversifying firms' and entries'

---

<sup>7</sup>The detailed breakdown of firm sales across sectors is available in the ESA survey, which is also maintained by the French Bureau of Statistics. In our main analysis, we define sectors according to the 5-digit code of the French Standard Industry Classification (SIC). We test the robustness of our main results by defining sectors at the 1-, 3-, or 5-digit levels of the French SIC.

<sup>8</sup>We develop a web-crawler that captures acquiring and target firms' names and addresses to link SDC Platinum and Zephyr deals to the French administrative data.

characteristics, to understand the interactions between human capital and other potential determinants of the decision to build or buy. First, we show that human capital plays a role in the decision to build or buy regardless of the size of the firm, the severity of financial constraints and whether it is publicly or privately owned. Second, we build two measures of distance between the firm's sector of origin and that of entry. The first measure is a product market distance measure that captures the complementarities between the firm's sector of origin and the sector of entry (e.g., production synergies, or a common customer base) Bloom, Schankerman and Van Reenen (2013). The second measure captures the physical distance between a firm and the geographical location of its entry. We find that both product market distance and physical distance to the market of entry increase a firm's likelihood to enter the market by acquiring an existing company. Moreover, the interaction of human capital with our distance measures shows that the magnitude of the human capital coefficient increases when a firm enters a product market that differs greatly from the firm's sector of origin. This finding suggests that internal resources matter more when the cost of building is likely to be high.

We check the robustness of our results by considering alternative measures of human capital and ruling out alternative mechanisms that could drive our results. First, we weight occupations by the number of employees in each occupation to account for the exact composition of the workforce, we focus on key occupations only, we exclude CEOs, whose wages may not only reflect productivity but also agency conflicts within the firm, and we modify the unit of observation from firms to plants. Second, we check that the scale of the new activity, as well as possible complementarities between human and physical capital (e.g., equipment, or machinery) do not fully confound the role of human capital in the decision to build or buy. On the one hand, we compare firms that operate in the same sector of origin, enter the same sector in the same year, and make similar sales. On the other hand, we compare firms that make similar capital expenditures when entering the new sector. Our point estimates remain nearly unchanged. Third, we check that firms that build and have a lower internal human capital adjust their workforce more following the entry. Indeed, firms with low human capital hire relatively more new workers, and especially in key occupations. Fourth, we exploit the heterogeneity in the workforce composition of subsidiaries within building firms, and we show that entry into the new sector is more often made through subsidiaries with higher internal human capital. This finding supports the view that firms minimize labor adjustment costs when contemplating entry in a new sector.

Furthermore, we attempt to mitigate concerns that firms likely take the decision to diversify *jointly* with the decision to build or buy in the sector of entry. Given that our sample consists of diversifying firms, self-selection could create a spurious

relationship between human capital and firm's decision to build or buy. We focus on two plausible scenarios after which a firm is likely to jointly select the sector of entry and the mode of entry. In the first scenario, we focus on firms that operate in a declining sector, and may thus be willing to shift their operations and reallocate workers to a better performing sector (Tate and Yang, 2016*b*; Baghai et al., 2018). We would then expect the firm to select a sector in which it can easily redeploy its workforce. In a second scenario, we focus on serial acquirers that are likely to always diversify by acquisition (Golubov, Yawson and Zhang, 2015). To test the first scenario, we exclude firms that shift a substantial part of their operations to another sector. For the second scenario, we exclude serial acquirers from the analysis. In both tests, we find our main results to hold, suggesting a limited role for selection issues.

Our second key finding is that firms are more likely to buy when it is costly to hire key workers on the external job market, i.e., when key worker occupations are in short supply. We test this second prediction from our model using occupation-level data on local labor market (LLM) tightness obtained from the French unemployment agency. The data report worker occupations that are in short supply across 350 LLMs. We document considerable geographic heterogeneity in worker availability across LLMs (Moretti, 2010). We create a measure of LLM tightness based on the weighted average human capital of occupations in short supply within the zip-code area where the firm enters. Our measure of LLM tightness takes higher values if key occupations for the sector of entry are in short supply in the LLM where the firm enters. We test whether the LLM tightness measure explains the decision to build or buy and interact it with our measure of human capital. We find that firms are more likely to buy when LLMs are tight, and that the relationship between human capital and the choice to build or buy is driven by the highest tercile of LLM tightness.

We then explore the performance implications of the choice to build or buy, and their relationship with labor market frictions. When LLMs are tight, hiring is costly. Therefore, after entry, we find that firms with less adapted human capital grow at a slower pace than firms with highly adapted human capital. In other words, firms with less adapted human capital sell less in the sector of entry. In addition, we find that this relationship is driven by sectoral entries that coincide with tight LLMs. Taken together, these findings suggest that labor market frictions are critical to understanding the role of human capital in diversification decisions.

Finally, we test whether firing costs affect firms' decision to build or buy. When a firm buys to enter a new sector, it has to incur the cost of restructuring the target (e.g., Dessaint, Golubov and Volpin, 2017). According to our model, the higher the restructuring cost, the less attractive is the option to buy. Furthermore, the more adapted an acquiring firm's internal human capital, the more workers must be laid off

after the acquisition because of a higher overlap of key worker occupations. Thus, higher firing cost makes the option to buy less attractive for firms with adapted internal human capital. In a first set of tests, we use the fraction of permanent to temporary workers in a firm's workforce to proxy for firing costs, hence the ease with which firms can restructure their workforce. However, firms endogenously decide whether to hire workers under permanent or temporary contracts. Therefore, in a second set of tests, we use the variation in the average length of local labor case settlements across jurisdictions to settle labor cases as a proxy for firing costs that is exogenous to firms (see, e.g., Fraisse, Kramarz and Prost, 2015). A higher length of labor case settlements is associated with higher firing costs. In both tests, we find a positive relationship between firing costs and firms' likelihood to grow by acquisition, though this relationship is not statistically significant. However, we find firing costs to interact significantly with our measure of human capital, suggesting that higher firing costs increase in the importance of existing human capital resources in firms' decision to build or buy.

**Related literature.** Our paper contributes to the rapidly growing literature on labor and corporate finance.<sup>9</sup> Specifically, our paper is related to a strand of papers that link the organization of firms and labor economics.<sup>10</sup> Closely related to our paper, Tate and Yang (2016*a*) predict diversified M&As by cross-industry labor flows under the condition that human capital is transferable across sectors. Ouimet and Zarutskie (2016) and Chen, Gao and Ma (2018) show that the desire to gain human capital is an important motive for corporate acquisitions. Lee, Mauer and Xu (2018) find that firms are more likely to merge and have better post-merger outcomes when the target firm has similar human capital. We make three contributions to this literature. First, we propose a firm-level measure of human capital that captures complementarities between firms and sectors. Second, we show that internal human capital resources predict how firms diversify. Third, our results emphasize the importance of labor shortages to understand the relationship between human capital and M&As. Overall, we show that employment composition shapes the boundaries of the firm.

Our paper also relates to the literature on the determinants of the decision to build or buy across sectors. Very few papers have jointly studied these alternative approaches to diversification.<sup>11</sup> McCardle and Viswanathan (1994) show that firms

---

<sup>9</sup>For instance, a strand of the literature has recently explored the implications of labor adjustment costs for corporate investment (e.g., Merz and Yashiv, 2007; Xu, 2018; Bai, Fairhurst and Serfling, 2018) and capital structure decisions (e.g., Matsa, 2010; Agrawal and Matsa, 2013; Simintzi, Vig and Volpin, 2014; Baghai et al., 2018; Serfling, 2016).

<sup>10</sup>Other papers study the consequences of organizational changes, such as M&As, on employment and wages (e.g., Lagaras, 2017; Ma, Ouimet and Simintzi, 2016).

<sup>11</sup>The literature on corporate diversification is large and has mainly focus on the sector choice

enter a new sector through acquisition when barriers to entry are high.<sup>12</sup> Phillips and Zhdanov (2013) show that in equilibrium, small firms invest in R&D, whereas large firms buy those small firms that have successfully innovated. Moreover, Bernard, Redding and Schott (2010) document that US multi-product firms often vary their product mix but infrequently do so through an acquisition (only 7% of cases). We contribute to this small strand of the literature by documenting that, at the scale of the French economy, more than 90% of corporate diversifications are made by firms that build from scratch using their preexisting resources rather than buying an incumbent in the sector of entry.

Finally, our paper adds to the literature on the theory of the firm. In the finance literature, the dominant view has been the “property rights” theory (Grossman and Hart, 1986; Hart and Moore, 1990; Rhodes-Kropf and Robinson, 2008), according to which value is created by combining complementary assets under the control of a single firm.<sup>13</sup> Another view from the early economics and strategy literatures has regained interest in trade and international economics. According to this “resource-based” view, the decision to grow depends on preexisting resources and transferable capabilities (Penrose, 1955; Chandler, 1992). Matsusaka (2001) theoretically shows that firms dynamically modify their portfolio of activities to match their organizational capabilities. Along these lines, Bernard et al. (2007) and Bernard et al. (2018) document that firms are much more likely to produce in certain pairs of industries. Boehm, Dhingra and Morrow (2019) take a step further and show that firms tend to co-produce in industries that require similar intermediate inputs. Our paper builds on this theory and focuses on another type of input: labor. We show that firms are more likely to diversify by building on their existing resources when complementarities exist between the firm’s internal human capital and the key skills needed to produce in the new sector. Our results imply that firms buy when they do not have adapted inputs in house, especially when these inputs are scarce on external markets.

---

in the decision to diversify. See Maksimovic and Phillips (2013) for a review of the literature.

<sup>12</sup>In the strategy literature, Yip (1982)’s empirical study supports McCardle and Viswanathan (1994)’s theoretical analysis and shows that firms are more likely to build in a sector with low barriers to entry. In the international economics literature, Nocke and Yeaple (2007) study the choice of entering a new country via foreign direct investment or acquiring an existing company.

<sup>13</sup>Rhodes-Kropf and Robinson (2008) show that the search for complementarities implies the existence of an assortative matching between acquirers and target firms, i.e., “like buys like”. An important strand of the empirical M&A literature finds evidence supporting this view: Mergers are more likely to occur and create more value when merging firms have similar human capital (Tate and Yang, 2016*b*; Lee, Mauer and Xu, 2018), sell similar products (Hoberg and Phillips, 2010), use similar technology (Bena and Li, 2014), or share similar corporate cultures (Li et al., 2018).

## 2.2. Theoretical framework: “build” or “buy”?

Why do some firms enter a new sector by acquiring an existing company (“buy”), while others do so using their existing resources (“build”)? We propose a model that predicts firms’ decision to build or buy based on the adaptability of their workforce to the new sector.

### 2.2.1. Basic framework

**Costs and profits.** To enter a new sector, firms must develop new productive capacities, that is, combine additional inputs to produce in the new sector. In the model, we assume that labor  $L$  is the only factor of production, so the production function is  $Y = \mathcal{L}$ .<sup>14</sup> Firms can select workers from three different pools. First, firms can reallocate workers from their *internal* labor market and have them produce in the new sector. Second, firms can hire new workers on the *external* labor market, that is, poach workers already employed or hire unemployed workers. Third, firms can *acquire* an existing company to have workers from the acquired firm produce in the new sector. The production function combines these three sources of labor with a constant elasticity (CES) of substitution:

$$\mathcal{L} = \left( \sum_i L_i^\gamma \right)^{\frac{1}{\gamma}}, \quad (2.1)$$

where  $0 < \gamma < 1$  is the elasticity of substitution across worker pools and  $i \in \{I, E, A\}$  denotes the different pools: internal, external, and acquired. Firms choose their mix of workers from the three pools to minimize their marginal cost of production  $c$ , defined as

$$c = \left( \sum_i c_i^{-\frac{\gamma}{1-\gamma}} \right)^{-\frac{1-\gamma}{\gamma}}, \quad (2.2)$$

where  $c_i$  denotes workers’ marginal cost of production in each pool  $i$ . We provide microeconomic foundations for each labor type’s marginal cost of production in Section 2.2.3.

We assume that firms engage in monopolistic competition for each variety of product and that consumer preferences exhibit a CES across products. Hence, firms’ profit is proportional to their marginal cost of production  $c$ :

$$\Pi = Kc^{-(\sigma-1)}, \quad (2.3)$$

where  $\sigma$  is the elasticity of substitution between goods within a sector,  $\sigma > 1$  and  $K$

---

<sup>14</sup>Note that in the empirical analysis, we consider the interactions between human and physical capital.

is a constant measuring firms' profitability.<sup>15</sup>

**Build.** If a firm chooses to enter the new sector by using its existing resources ("build"), it can choose to combine workers from its internal labor market ( $L_I$ ) and the external labor market ( $L_E$ ). We assume that building does not have any impact on firms' existing business lines of production. This assumption is realistic if the introduction of new tasks does not disrupt existing tasks, for instance, if internal workers are shifted away from tasks that are either completed or sufficiently automated.

$$\Pi^{Build} = K \left( c_I^{-\frac{\gamma}{1-\gamma}} + c_E^{-\frac{\gamma}{1-\gamma}} \right)^{\frac{(1-\gamma)(\sigma-1)}{\gamma}}. \quad (2.4)$$

Equation (2.1) implies that the three pools of workers are combined according to the CES parameter  $\gamma$ . Therefore, in addition to consumers' elasticity of substitution between products  $\sigma$ , a firm's monopolistic profit in (2.4) depends on the marginal costs of production of the different worker pools, aggregated according to a function of  $\gamma$ .

**Buy.** If a firm chooses to enter a new sector by acquiring an existing company ("buy"), it accesses the target's pool of workers ( $L_A$ ), in addition to its internal labor market ( $L_I$ ) and the external labor market ( $L_E$ ). Accessing target workers is desirable because target firms' human capital is likely to be more adapted to the sector of entry, as discussed in Section 2.2.3. However, we assume that incorporating the target's full pool of workers when acquiring a firm's internal labor market leads to post-merger restructuring costs. We model these costs as a fixed cost  $F$ . In Appendix 2.11.1, we propose a microfoundation of the fixed cost  $F$  as the costs of restructuring the workforce of the acquired firm. When firms buy, their profit can be written as:

$$\Pi^{Buy} = K \left( c_I^{-\frac{\gamma}{1-\gamma}} + c_E^{-\frac{\gamma}{1-\gamma}} + c_A^{-\frac{\gamma}{1-\gamma}} \right)^{\frac{(1-\gamma)(\sigma-1)}{\gamma}} - F. \quad (2.5)$$

The acquiring firm's access to the target firm's worker pool increases the variable part of (2.5). When a firm buys, we can theoretically show that it reallocates fewer workers from its internal labor market and hires fewer workers on the external labor market.

---

<sup>15</sup>In standard monopolistic competition models, this constant depends on  $\sigma$  and the share of consumer income devoted to a given product. In turn, this share depends on the price of a given product relative to the aggregate price index. The only relevant detail for our analysis is that this constant does not depend on the marginal cost, as is the case in standard monopolistic competition models.

Moreover, firms systematically lay off some of the target's workers. These findings provide an interpretation for the fixed cost  $F$  as the post-merger restructuring costs, which is in line with LeeMauer, who show that laying off duplicate workers is a source of economies of scale in mergers.

### 2.2.2. Testable predictions: build or buy?

When entering a new sector, the firm compares the profit from building on its existing resources ("build") with the realized profit from acquiring an existing company ("buy"). We show in Appendix 2.11.1 that if the different pools of workers are sufficiently substitutable ( $\gamma > (\sigma - 1)/\sigma$ ) and  $F$  is low enough, the model implies that there exists a unique threshold  $c_I^* > 0$  such that if  $c_I > c_I^*$ , we have  $\Pi^{Buy} > \Pi^{Build}$ . Otherwise, if  $c_I < c_I^*$ , we have  $\Pi^{Buy} > \Pi^{Build}$ . In this way, we obtain prediction 2.2.2.

Firms optimally choose to build when their marginal cost of production is low enough, that is, if their existing workforce is more adapted to operate into the sector of entry.

As we show in Appendix 2.11.1, the model also implies  $\frac{\partial c_I^*}{\partial c_E} < 0$ , from which prediction 2.2.2 follows.

Firms are more likely to buy when key workers for the sector of entry are in short supply in the external labor market.

Prediction 2.2.2 implies that prediction 2.2.2 is stronger when key workers for the sector of entry are in short supply in the external labor market.

### 2.2.3. Micro-foundations of the labor cost

To take predictions 2.2.2 and 2.2.2 to the data, we need a micro-foundation for the labor costs  $c_i$ . In this section, we propose one, based on Cheng and Morrow (2018), in which labor costs depend on the availability and efficiency of workers in each pool.

**Worker occupations.** We assume that there are different types of workers. Each worker type is employed in a given occupation. Each worker occupation, denoted  $o \in \mathcal{O} = \{1, \dots, O\}$ , is available in quantity  $(a_{i1}, \dots, a_{iO})$ . Importantly, worker availability differs across worker pools  $i \in \{I, E, A\}$ : The number of workers in each occupation is different in the acquiring firm's internal labor market, in the external labor market, and in the target's internal labor market. Worker characteristics differ across occupations: Workers' wages  $(w_1, \dots, w_O)$  and efficiency  $(m_1, \dots, m_O)$  differ across occupations but not across workers within an occupation.<sup>16</sup>

---

<sup>16</sup>We, therefore, drop the subscript  $i$  relative to wages and efficiency.

**Recruiting process.** To recruit workers, the firm conducts interviews with several workers from each pool  $i \in \{I, E, A\}$ . The match between a firm and a worker is assumed to be of random quality  $h \geq 1$  and to follow a Pareto distribution with cumulative density function  $\Psi(h) = 1 - h^{-k}$  with  $k > 1$ . This distribution is assumed to be equal across worker pools; that is, firms' expected match quality is the same regardless of whether workers come from inside the firm (i.e., the internal labor market) or from outside (i.e., the external labor market or from the target firm's workforce). This assumption holds as long as workers' new tasks in the sector of entry are sufficiently different from the tasks to which they were previously assigned. In this case, there is no reason for firms to have different expected match quality from workers from different pools.<sup>17</sup> Firms observe the match quality during the interview.

We assume that conducting interviews is labor-intensive and costs  $f$  per interview. This assumption implies that the degree of information asymmetry between firms and workers does not vary across the three worker pools. There are two reasons that we assume the cost  $f$  to be the same across the different pools of workers. First, this assumption simplifies the theoretical analysis and does not change the predictions. Second, empirically, we cannot observe the cost of interviews and hiring costs across firms or differentiate these costs across worker pools.

The firm selects workers to interview by fixing a match quality threshold  $\underline{h}_{io}$  below which they do not hire workers they interview. After the interviews, the firm hires a total number  $N_i a_{io} \underline{h}_{io}^{-k}$  of workers in each occupation  $o \in \mathcal{O}$ . The total labor cost of conducting interviews is then  $f N_i c_i$ .

**Costs and production function.** The firm's total labor cost  $C_i$  in each pool of workers  $i$  sums to the cost of conducting interviews  $f N_i c_i$  and the wages of selected workers:

$$C_i = N_i \left( \sum_o a_{io} w_o \underline{h}_{io}^{-k} + f c_i \right). \quad (2.6)$$

For a given number of interviews  $N_i$  and a quality threshold  $\underline{h}_{io}$ , a firm can estimate its workforce's expected level of human capital. Each worker occupation  $o \in \mathcal{O}$  is associated with an expected level of human capital equal to  $H_{io}$ .<sup>18</sup> Within each pool of workers, occupation's expected human capital represents the input in

<sup>17</sup>In the empirical analysis, we control for the product market distance between firms' existing business lines and the new sector of production. We find that the data continue to support our model's predictions. Moreover, we find stronger support for our model's predictions when the new sector is distant from firms' existing sectors of activity, that is, when it is most likely that  $k$  is the same across worker pools.

<sup>18</sup>The match quality follows a Pareto distribution, which implies a simple formula for each worker occupation's expected human capital:  $H_{io} \equiv N_i a_{io} m_o \int_{\underline{h}_{io}}^{\infty} h d\Psi(h) = \frac{N_i a_{io} m_o k \underline{h}_{io}^{1-k}}{k-1}$ .

the firm's production function (2.1) as follows:

$$L_i = (H_{i1}^\theta + \dots + H_{iO}^\theta)^{1/\theta}, \quad (2.7)$$

where  $\theta < 1$  is the elasticity of substitution between workers' human capital across different occupations. We assume the elasticity to be constant across worker pools  $i \in \{I, E, A\}$ . This assumption means that the substitutability across occupations is determined by the nature of tasks involved in the production process.

Our model requires that both the interview cost *and* the match quality distribution be constant across worker pools. As a consequence, workers from the internal, external, and the acquired firm's labor markets are perfectly substitutable. Hence, the model does not allow for firm-specific organizational capital or firm-specific human capital. The variation in labor costs across worker pools comes from the different vectors of occupation availability  $(a_{i1}, \dots, a_{iO})$ .

**Cost minimization.** The firm minimizes the total labor cost  $C_i$  in each worker pool (2.6) by choosing the number of interviews  $N_i$  to conduct and the occupation-specific match quality threshold  $\underline{h}_{io}$ . Note that the firm chooses  $\underline{h}_{io}$  and  $N_i$  by trading off the quality of the hired workers and the search costs associated with recruiting process. On the one hand, hiring a large number  $N_i$  of workers enables firms to select the best matches by choosing high values for  $\underline{h}_{io}$ . On the other hand, the firm saves on search costs  $f$  by choosing a smaller number of workers  $N_i$  and low values of  $\underline{h}_{io}$ . However, the firm takes as given the expected human capital supplied by its workers  $H_{io}$  and the production function (2.7).

We show in Appendix 2.11.1 that this constrained minimization problem results in the following marginal labor cost:

$$c_i = \left( \sum_{o \in O} \left( \frac{a_{io} m_o^k w_o^{1-k}}{f(k-1)} \right)^{\frac{\theta}{\beta}} \right)^{\frac{\beta}{\theta(1-k)}}, \quad (2.8)$$

where  $\beta \equiv \theta + \theta(1-k)$ .<sup>19</sup> Equation (2.8) implies that the marginal labor cost in worker pool  $i \in \{I, E, A\}$  increases with the interview cost  $f$ . Indeed, higher interview costs reduce the optimal number of interviews. Matches' average quality decreases, and, in turn, the human capital supplied by each worker in an occupation is lower. This expression highlights that the labor cost is determined by the byproduct of workers' efficiency and availability. It is large when occupations that are efficient in the new sector (high  $m_o$ ) are in scarce supply in the worker pool (low  $a_{io}$ ).

---

<sup>19</sup>The assumption that  $k > 1$  implies that  $\beta > 0$ , in which case it is optimal for firms to hire workers in every occupation (Cheng and Morrow, 2018).

**Wage-bill share of each occupation.** We denote by  $A_{io}$  the total number of workers in occupation  $o$  hired to produce one unit of output. We show in Appendix 2.11.1 that the share of the wage bill that goes to a given occupation  $o$  can be expressed as:

$$\frac{w_o A_{io}}{\sum_o w_o A_{io}} = \frac{(a_{io} w_o^{1-k} m_o^k)^{\frac{\theta}{\beta}}}{\sum_o (a_{io} w_o^{1-k} m_o^k)^{\frac{\theta}{\beta}}}. \quad (2.9)$$

Equation (2.9) implies that within a worker pool  $i \in \{I, E, A\}$ , the share of a firm's wage bill that goes to workers in a given occupation depends on (i) occupation-specific wages  $w_o$ , (ii) the availability  $a_{io}$  of workers in that occupation, and (iii) workers' occupation-specific efficiency  $m_o$ . The interpretation is that within a given sector, some occupations receive relatively higher wages than other occupations when they are relatively more efficient at producing in this sector.

## 2.3. Empirical strategy

To take predictions 2.2.2 and 2.2.2 to the data, we need to estimate the marginal labor cost (2.8) for workers from the internal labor market, i.e.,  $c_I$ . In Section 2.4, we explain that neither the availability  $a_{Io}$  of internal labor market workers in each occupation nor the occupation-specific efficiency  $m_o$  are observable in the data. However, based on our model, we propose a method to estimate  $c_I$  using administrative worker-level occupation data.

### 2.3.1. Occupation-specific human capital

Equation (2.9) allows us to estimate the average availability  $a_{Io}$  of workers in occupation  $o$  and the occupation-specific efficiency  $m_o$ . We use the subscripts  $f$ ,  $o$ ,  $n$ , and  $t$ , for firm, occupation, the sector of the firm, and time, respectively. At the firm level, we denote by  $\text{Share}_{f,o,n,t}$  the share of the wage bill that goes to occupation  $o$  in firm  $f$  operating in sector  $n$ , i.e., the ratio given by Equation (2.9). Rewriting worker availability  $a_{f,o,n,t}$  as a deviation from the sectoral average  $\bar{a}_{o,n,t}$ , we have  $a_{f,o,n,t} = \bar{a}_{o,n,t} \cdot \tilde{a}_{f,o,n,t}$ . Then, taking the logarithm of Equation (2.9), we obtain

$$\log(\text{Share}_{f,o,n,t}) = \frac{\theta}{\beta} \log(\bar{a}_{o,n,t} m_{o,n,t}^k w_{o,n,t}^{1-k}) - \log\left(\sum_{o' \in \mathcal{O}_{gt}} (a_{g,o',t} m_{o',n,t}^k w_{o',n,t}^{1-k})^{\frac{\theta}{\beta}}\right) + \frac{\theta}{\beta} \log(\tilde{a}_{f,o,n,t}), \quad (2.10)$$

where  $\mathcal{O}_{ft}$  is the set of occupations observed in firm  $f$  at time  $t$ . Equation (2.10) provides a decomposition of  $\text{Share}_{f,o,n,t}$ , which can be estimated as the following fixed effects regression:

$$\log(\text{Share}_{f,o,n,t}) = \mu_{o,n,t} + \nu_{f,t} + \epsilon_{f,o,n,t}, \quad (2.11)$$

where  $\mu_{o,n,t}$  is an occupation  $\times$  sector  $\times$  time fixed effect capturing the average wage share that goes to occupation  $o$  at the level of sector  $n$ .  $\nu_{f,t}$  is a firm  $\times$  time fixed effect, and  $\epsilon_{f,o,t}$  is an error term capturing the deviation of occupation  $o$ 's share in the firm's wage bill from the sectoral average.

The estimation of  $\mu_{o,n,t}$  requires firms to hire more than one type of occupation each year and every occupation to be present in more than one firm in a given sector in each year.<sup>20</sup> In Table 2.A1, we report the explanatory power of our first stage Equation (2.11). Column 1 shows our first stage as used in the rest of the paper. The dependent variable is the (log) share of a firm's wage bill that goes to a specific occupation and we regress it on Occupation-Sector-Year FE *and* Firm-Year FE. In columns 2-6, we look at different combinations of fixed effects. We find that 72.3% of the within-firm variation in the log-share allocated to a specific occupation is explained by Occupation-Sector-Year FE and Firm-Year FE. Occupation-Sector-Year FE alone explain 53% of the variation (column 2), and Firm-Year FE alone explain 51% of the variation, suggesting that the combination of the two sets of fixed effects significantly improves the percentage of the within-firm occupation log-share's variation explained by our first stage displayed in column 1.

We use the estimated values, denoted  $\hat{\mu}_{o,n,t}$ , to construct our main explanatory variable below (Section 2.3.2). We interpret them as a score reflecting the (wage- and availability-adjusted) human capital of a given occupation at the sectoral level. The higher the fixed effect of a given occupation in a sector is, the larger the score for this occupation in this sector.<sup>21</sup>

### 2.3.2. Firm's internal human capital

We rely on the occupation scores to analyze the role of human capital in the decision to build or buy. We use the subscripts  $g$ ,  $o$ ,  $n$ , and  $t$ , for the diversifying firm, worker occupation, sector of entry, and time, respectively. The marginal labor cost (2.8) can then be rewritten as  $c_{I,g,n,t} = (f(k-1))^{\frac{1}{k-1}} (\psi_{I,g,n,t-1})^{\frac{\theta}{\theta(1-k)}}$ , where

$$\psi_{I,g,n,t} = \sum_o (a_{I,g,o,n,t} w_{g,o,n,t}^{1-k} m_{g,o,n,t}^k)^{\frac{\theta}{\beta}}. \quad (2.12)$$

We interpret  $\psi_{I,g,n,t}$  as a proxy for the human capital of workers from a firm's internal labor market. Indeed,  $c_{I,g,n,t}$  is a decreasing function of  $\psi_{I,g,n,t}$ , that is, the marginal

---

<sup>20</sup>To ensure that these two conditions are met, we exclude firms that hire only one type of occupation and employ fewer than 20 workers from the sample. In addition, in the empirical analysis, we exclude occupation-sector-year triplets with fewer than 10 firms to obtain more precise estimates.

<sup>21</sup>As an illustration, Table 2.A2 displays the five occupations with the largest fixed effects in the sectors of pharmaceutical preparations, IT consultancy activities and the manufacture of motor vehicles in 2013. The selected occupations indeed seem to play an important role in their sectors.

cost of labor of operating in sector  $n$  is lower when the human capital of firms' existing workers is high for that sector.

Rewriting  $\psi_{I,g,n,t}$  in terms of the occupation  $\times$  sector  $\times$  time fixed effect  $\mu_{o,n,t}$  and the firm-level deviation from the sectoral average availability of workers in an occupation,  $\tilde{a}_{g,o,t}$ , we have

$$\psi_{I,g,n,t} = \sum_o (\tilde{a}_{g,o,n,t})^{\frac{\theta}{\beta}} \exp(\mu_{o,n,t}). \quad (2.13)$$

Predictions 2.2.2 and 2.2.2 consist of predicting firms' choice to build or buy, based on their existing workforce composition. Therefore, we need to empirically estimate firm  $g$ 's human capital  $\psi_{I,g,n,t-1}$  one year before entering the new sector.

The key challenge to estimating (2.13) is that we do not observe firms' internal reallocation of existing workers in the sector of entry; that is, we do not observe  $\tilde{a}_{g,o,n,t}$ . Therefore, as soon as occupation  $o$  is present in  $g$ 's existing workforce at  $t - 1$ , we assume that  $\tilde{a}_{g,o,n,t} = 1$ , meaning that worker availability in the internal labor market is equal to the average availability in the new sector. Otherwise, if  $a_{g,o,n,t} = 0$ , we assume that  $\tilde{a}_{g,o,n,t} = 0$ .<sup>22</sup> We obtain our main independent variable, which is an empirical estimation of the value of (2.13) at  $t - 1$ :

$$\text{Internal Human Capital}_{g,n,t-1} = \frac{1}{\#O_{g,t-1}} \cdot \sum_{o \in O_{g,t-1}} \exp(\hat{\mu}_{o,n,t-1}), \quad (2.14)$$

where  $O_{g,t-1}$  is the set of occupations present in firm  $g$ 's internal labor market prior to entering the new sector  $n$  (i.e., occupations for which  $a_{g,o,n,t-1} > 0$ ) and the values of  $\hat{\mu}_{o,n,t-1}$  come from the estimation of Equation (2.11).

Internal Human Capital $_{g,n,t-1}$  is a measure of the human capital of firm  $g$ 's existing workforce for the sector of entry  $n$ . It is defined as the average value of the (exponentiated) occupation  $\times$  sector  $\times$  year fixed effects of the occupations present in the workforce of the firm prior to entry, given by (2.11). We use the average value instead of the sum to avoid human capital being mechanically larger for firms employing workers in more occupations.<sup>23</sup>

Internal Human Capital $_{g,n,t-1}$  takes a high value when occupations with a high (wage-adjusted) efficiency  $m_{o,n,t-1}$  for the sector of entry are already present in a firm's internal labor market prior to entry. We interpret it as a measure of whether

<sup>22</sup>In the empirical analysis, we test several alternative assumptions as robustness checks (see Section 2.6.1). In particular, we use the fraction of each occupation as weights in the computation of Internal Human Capital $_{g,n,t-1}$ . Our results are robust to alternative assumptions in our estimation of Equation (2.13).

<sup>23</sup>We also control for firm size using the number of employed workers. In unreported regressions, we control for the number of occupations to closely follow the definition of  $\psi_{I,g,n,t-1}$ . Our results are robust to this change.

a firm’s workforce is adapted for a given sector of entry. In the remainder of the paper, we refer to this measure as “human capital” or “internal human capital” without explicitly mentioning that it is specific to a given sector of entry.

### 2.3.3. Empirical model

Prediction 2.2.2 states that firms are more likely to “buy”, as opposed to “build”, when their existing internal human capital is not adapted to the market of entry. This pattern is more pronounced when key workers for the sector of entry are difficult to hire in the external labor market (prediction 2).

We test these predictions by analyzing the link between the type of entry and the constructed measure of internal human capital. The dependent variable is  $1(\text{Buy})_{g,n,o,t}$ , a dummy equal to one if firm  $f$  enters a new sector  $n$  through an acquisition (“buy”) and zero if it enters the new sector by building on its existing resources (“build”).  $o$  indicates firm  $f$ ’s main original sector of activity, i.e., the sector in which the firm realizes the largest share of its sales prior to entry. Our baseline empirical model is as follows:

$$\mathbb{1}(\text{Buy})_{g,n,o,t} = \lambda_{n,o,t} + \delta \text{Internal Human Capital}_{g,n,t-1} + \beta X_{g,n,o,t-1} + \eta_{g,n,o,t} \quad (2.15)$$

This presence of origin  $\times$  entry  $\times$  time fixed effects implies that the role of internal human capital is identified by comparing the diversification strategy of firms operating in the same sector of origin and entering in the same new sector. All unobservable time-varying synergies and complementarities between sectors are therefore captured by the fixed effects  $\lambda_{n,o,t}$ . In addition, this specification also controls for unobservable factors related to the sector of entry (e.g., fixed costs of entry, barriers to entry. See McCardle and Viswanathan (1994)) or the sector of origin (e.g., ability to collateralize assets to access external finance. See Rajan and Zingales (1998)).

The vector  $X_{g,n,o,t-1}$  includes other firm characteristics that may influence the decision to build or buy. At the firm level, such determinants are firm size, cash holding, tangibility, labor productivity and product market relatedness (i.e., the distance to the sector of entry based on sales information). We also control for the variation in labor market tightness across geographical zones and sectors.

By construction, internal human capital $_{g,n,t-1}$  introduces a measurement error term that generates a correlation between residuals  $\eta_{g,n,o,t}$  at the level of the sector of entry. It might also be more or less precisely estimated depending on the sector of origin. We therefore double-cluster standard errors to control for correlations within the sector of entry and within the sector of origin. Finally, we standardize the internal human capital $_{g,n,t-1}$  variable in the rest of the analysis to interpret

regression coefficients in standard deviation units.

## 2.4. Data and summary statistics

### 2.4.1. Data sources

#### Firm data

Our primary source of data is French administrative data provided by the French Bureau of Statistics (INSEE). Our definition of a firm includes the parent company and the majorly owned subsidiaries (more than 50% of the shares), which we identify using a dataset containing ownership links (*Enquête sur les Liaisons financières entre sociétés*, LIFI). We recover firms' main sector of activity, balance-sheet information and income statements from the tax files (*Bénéfices Industriels et Commerciaux* and *Bénéfices Non-commerciaux*). We use the SIRENE registry to obtain the geographical location of the different plants of each firm. We then consolidate all variables to obtain observations at the parent firm-level.<sup>24</sup>

To identify diversified entries into new sectors, we rely on the subsidiaries' breakdown of sales by sector (*Enquête Annuelle de Production*). This dataset records the detailed amount of sales realized by subsidiaries in every sector and year. The survey is exhaustive for subsidiaries with at least 20 employees and randomly includes smaller subsidiaries, such that the survey covers at least 85% of sales realized within a given sector. For instance, subsidiaries included in the survey cover 96% of sales in the manufacturing sector. For smaller subsidiaries for which sales breakdowns are not available in the survey, we make the assumption that these subsidiaries sell only in their main sector of activity, and these data are retrieved from the tax files. We exclude entries occurring in 2008 because the methodology of the survey changed that year.

Sectors are defined by the French Standard Industry Classification (SIC) (*Nomenclature des activités Françaises, NAF*), which is equivalent to the US SIC. We define sectors by 5-digit French SIC codes. The dataset includes 732 different sectors.<sup>25</sup>

---

<sup>24</sup>Note that we exclude firms in the financial, agricultural and public sectors because they use different accounting systems, which limits the comparability and relevance of standard variables across sectors.

<sup>25</sup>Note that we would not be able to replicate Hoberg2010 product market distance with our data. This measure is based on textual analysis on 10k filings of U.S. firms, information that is not available in France and especially for both public and private firms.

### Merger and acquisition deals

We merge the French administrative datasets with a dataset of M&A deals retrieved from SDC Platinum and Bureau van Dijk Zephyr.<sup>26</sup> We collect all deals between January 2003 and December 2014 that involve a French acquirer and a French target. We exclude leveraged buyouts and private equity deals from the sample. We focus on deals in which the acquirer owns less than 50% of the target shares before the acquisition date and more than 50.1% after to identify changes in majority ownership between operating companies.

SDC Platinum and Bureau van Dijk's Zephyr do not provide French firm standardized identifying numbers (SIREN). We proceed in several steps to retrieve the unique firm identifiers. First, we use tickers (available only for publicly traded firms) and the Bureau van Dijk identifiers (available only for Zephyr deals) to recover a fraction of the firm identifiers. Second, we build a Python webcrawler on two websites, which takes as inputs a firm's name and address.<sup>27</sup> (i) [www.bodacc.fr](http://www.bodacc.fr) (*Bulletin Officiel des Annonces Civiles et Commerciales*), which is a governmental website that reports official notifications involving French companies since 2003, and (ii) [www.societe.com](http://www.societe.com), which is a commercial website that aggregates information about French companies from various sources (mostly from the French Bureau of Statistics and *Bodacc.fr*). Both websites are supposed to cover the universe of French firms. Third, after running the web-crawler, we drop companies for which the address, city and zip code are missing because we cannot identify with certainty the corresponding company identifier among several matches. We retain only observations for which the Jaró-Winkler string distance to the original name is below a certain threshold.<sup>28</sup> We retain observations with a distance above 0.8. Fourth, we manually check the resulting matches. Our final sample contains 7,303 deals from 2003 to 2014. To the best of our knowledge, this is the most comprehensive M&A dataset available to date for the French economy.

### Worker-level occupation worker data

We use the French matched employer-employee administrative dataset (*Déclarations Annuelles des Données Sociales*, DADS) to construct the measure of internal human capital. Firms are required by law to report every year detailed information about

---

<sup>26</sup>Note that an ownership change in the ownership links dataset cannot be directly used to identify M&As deals. Some ownership changes correspond to new entries in the database that do not necessarily correspond to new ownership links.

<sup>27</sup>The webcrawler builds on the Python packages Selenium and BeautifulSoup.

<sup>28</sup>The Jaró-Winkler distance measures the number of characters in common between strings under the assumption that differences near the start of the string are more significant than differences near the end of the string.

their workers when filing payroll taxes.<sup>29</sup> The employer must report the type of contract, gross and net wages, the number of hours worked and an occupation code for each worker. The data set also indicates whether the worker is employed with a permanent contract (*Contrat à durée indéterminé*, or CDI) or not. Note that before 2003, occupations are often missing from the dataset. This is why we begin the empirical analysis at this date. Occupations are reported as 4-digit codes. The French nomenclature of occupations (*Nomenclatures des professions et catégories socio-professionnelles des emplois salariés des employeurs privés et publics*, PCS-ESE) consists of 414 different occupations, including, for instance, 28 different types of engineers (e.g., logistics, IT, electrical, or mechanical).

### **Labor market frictions and firing costs**

We measure hiring frictions at the level of the local labor market (LLM) using data from the French national unemployment agency (*Pôle emploi*). The unemployment agency lists job vacancies, helps unemployed people find jobs and produces national and local unemployment statistics, which we use in this paper. In particular, the unemployment agency tracks occupations in short supply in 350 different LLMs starting in 2010. Occupations are flagged as being in short supply when (i) job offers exceed job applications and (ii) surveyed employers anticipate that they will not fill in a job. Figure 2.3 maps the number of occupations in short supply by LLM in 2013.<sup>30</sup> Darker shades of blue indicate a higher degree of shortage in the LLM. Interestingly, we see that labor market tightness is not systematically related to population density, as tight LLMs can be observed both in urban and rural areas.

To measure firing costs, we use data about 2010 local labor courts from the French Ministry of Justice website that deal with labor disputes between firms and workers (“Prud’hommes”). Elected judges representing employers and employees seat in the local labor courts. The local court in which the labor case is settled is determined by the firm’s location. Due to a judicial reform, the number of local labor courts decreased from 271 to 210 in 2008. Therefore, to ensure a constant coverage over time, we aggregate LLM-level observations at a more aggregate level of the judicial map, called a “jurisdiction”. There are 140 jurisdictions over our sample period.

To proxy for firing costs, we use variation across jurisdictions in the average length of labor case settlements. The implicit assumption here is that the length of local labor case settlements reflects a high firing cost for firms, e.g. because the length is due to frequent workers litigations (Fraissee, Kramarz and Prost, 2015).

---

<sup>29</sup>Note that reporting of the occupation code is required for firms that employed at least 20 employees in a given year and optional for firms below the threshold.

<sup>30</sup>Appendix Table 2.A4 lists occupations that are in short supply in 2013.

## 2.4.2. Main variables

### Type of entry

The main dependent variable  $\mathbb{1}(\text{Buy})_{g,n,t}$  is a dummy variable equal to one if the entry of firm  $f$  in sector  $n$  at time  $t$  is made through an acquisition (“buy”) and equal to zero if the entry is made internally (“build”). We consider a firm to have entered a new sector if (i) at least one of its subsidiaries begins selling in that sector and (ii) none of the other subsidiaries already operates in the sector.

A firm enters a new sector by acquisition if the entity that reports sales in the new sector (“entering subsidiary”) becomes one of the acquirer’s subsidiaries after the M&A. By contrast, a firm builds if the entering subsidiary had already been controlled by the firm<sup>31</sup>.

We also look at whether the entry in the new sector is associated with an entry in a new geographical zone. We use the “region” as definition of the geographical zone (France had 25 regions over the sample period). A firm starts operating in a new geographical zone if either (i) the entering subsidiary is created at time  $t$  and located in a new region, (ii) the entering subsidiary opens a plant in a new region at time  $t$  or (iii) the entering subsidiary is acquired at time  $t$  and operates in a region in which the firm was not present at  $t - 1$ .

### Distance between sector of origin and sector of entry

To study how the complementarities between the sector of entry and the sector of origin interact with internal human capital, we construct a product market distance measure based on Bloom, Schankerman and Van Reenen (2013). The firm-level variable measures the distance between the firm’s sectoral repartition of sales (“sector portfolio”) prior to entry to the sector portfolios of firms already operating in the sector of entry. The idea of the metric is that if the sector portfolio of the entering firm is unusual compared to firms already present in the sector of entry, then the firm is “distant” to the sector of entry - otherwise it is “close”.

For a firm  $f$  entering a new sector at time  $t$ , we denote by  $S_g = (S_g^1, \dots, S_g^N)$  the vector of sales at time  $t - 1$  broken down by sectors ( $n = 1, \dots, N$ ).  $S_g^{-n}$  is the vector of sales excluding sales in sector of entry  $n$ .<sup>32</sup> For a given sector of entry  $n$ , we define the distance  $d_{g,h}^n$  between the firm and any firm  $h$  already operating in the new sector  $n$ , as (one minus) the uncentered Pearson correlation between vectors

---

<sup>31</sup>See Appendix Figure 2.A1 for an illustration of the build and buy entries.

<sup>32</sup>We exclude sales in sector  $n$  because a firm that enters sector  $n$  at time  $t$  necessarily reports zero sales in  $n$  at time  $t - 1$ .

$S_g^{-n}$  and  $S_h^{-n}$ :

$$d_{g,h}^n = 1 - \frac{(S_g^{-n} \cdot S_h^{-n})}{\sqrt{(S_g^{-n} \cdot S_g^{-n})} \sqrt{(S_h^{-n} \cdot S_h^{-n})}}$$

Finally, we define the product market distance of firm  $f$  to sector  $n$  as the weighted average of the distance  $d_{g,h}^n$  for all firms  $h$  operating in sector  $n$  at  $t - 1$ :

$$\text{Product Market Distance}_{g,n} = \frac{\sum_h \omega_h^n d_{g,h}^n}{\sum_h \omega_h^n}$$

where the weights  $\omega_h^n$  are given by the share of sales realized by firm  $h$  in sector  $n$  at  $t - 1$ :  $\omega_h^n = \frac{\text{Sales of firm } h \text{ in sector } n}{\text{Sales of firm } h}$ . The weights ensure that the distance between the diversifying firm and firm  $h$  matters more if firm  $h$  realizes a large part of its sales in sector  $n$ .

### Local labor market tightness and length of local labor case settlements

We use the list of occupations in short supply by LLM to build a time-, sector- and geographic-level measure of labor market tightness. We define the *Local Labor Tightness* $_{n,z,t-1}$  variable as the average of the (exponentiated) occupation  $\times$  sector  $\times$  year fixed effects of the occupations in short supply in LLM  $z$ :

$$\text{Local Labor Tightness}_{n,z,t-1} = \frac{1}{N.L_{n,t-1}} \sum_{l \in L} \mathbb{1}(l \text{ in short supply in LLM } z) \times \exp(\hat{\mu}_{o,n,t})$$

where  $N.L_{n,t-1}$  is the number of occupations in sector  $n$  at time  $t - 1$ .<sup>33</sup> This sector-specific measure of LLM tightness takes high values if there are occupations in short supply in LLM  $z$  at time  $t - 1$  that are key for firms already operating in sector  $n$ . Figures 2.4a and 2.4b plot the geographical distribution of the variable for manufactures of pharmaceutical preparations and motor vehicles. The graphs show that firms face, on average, more difficulties finding key workers in the second sector than in the first sector. Moreover, the measure appears to vary significantly across both LLMs and sectors.

We then use the average length of labor case settlements as a time- and geographic-level proxy for firing costs. The average length of labor procedures is 12.5 months (standard deviation: 3.6). We rely on the location of the entering subsidiary to determine the intensity of local labor market frictions faced by the firm.

---

<sup>33</sup>In a robustness check, we directly use the percentage of jobs in short supply as a proxy for LLM tightness:

$$\text{Local Labor Tightness}_{n,z,t-1} = \frac{1}{N.L_{n,t-1}} \sum_{l \in L} \mathbb{1}(l \text{ in short supply in LLM } z).$$

Our results are robust to this alternative specification.

### 2.4.3. Summary statistics

**Build or buy?** Panel A of Table 2.1 and Figure 2.1a present the evolution of the proportion of build and buy entries between 2004 and 2013 (excluding 2008). While at the beginning of the period, approximately 1.8% of entries are made by acquisition, this figure increases over the sample period to reach 2.49% of total entries in 2014. However, buy entries are, on average, larger than build entries. Although sales are larger for build entries compared to buy entries in aggregate (Figure 2.1b), when weighting by entry sales, buy entries go up to an average 8.5% of total entries between 2004 and 2013 (Panel B of Table 2.1).

We then examine how these figures vary with the definition of sector in Panel B of Table 2.1. The proportion of buy entries remains stable at 1.7-1.9% whether we define a sector using the 1-digit or 5-digit code of the French SIC (our baseline). This finding suggests that firms do not tend to enter more by acquisition in sectors that are classified as similar in the French administrative data.<sup>34</sup> However, we find that build entries tend to be larger in sectors that are close to the sector of origin. Indeed, entries by acquisition represent 15% of entry sales when we use the 1-digit classification level but only 8% when we rely on the most detailed definition of sector (5-digit code).

[Insert [intext]Table 2.1 here]

**Internal human capital.** Figure 2.2 plots the probability density functions of firms' human capital by type of entry. The variable Internal Human Capital<sub>*g,n,t-1*</sub> is normalized to have a mean of zero and a standard deviation of one. On average, the human capital of firms that enter a new sector by acquisition is less adapted to the sector of entry than that of firms that choose to build.

[Insert [fig]poly]Figure 2.2 here]

**Other variables.** Table 2.2 Panel A reports summary statistics on the different control variables we include in our baseline specification (see Section 2.3.3). Consistent with Table 2.1, the average number in the panel of firms diversifying through an acquisition is approximately 2%. Firms report on average €2.82 million sales in the sector of entry in the first year, with a very large dispersion around the mean. The same year, we find that they invest on average €960,000. Firms enter industries that are vertically integrated into their original industry in 67% of the cases and usually in sectors that are close to their sector of origin. They employ approximately 630

---

<sup>34</sup>In the empirical analysis, we control for product market distance using our measure from Section 2.4.2 instead of the French SIC codes and find that firms tend to buy more frequently in more distant product markets (2.3). Hoberg and Phillips (2016) discuss several improvements over the core method underlying static sector classifications.

workers, produce approximately €50,000 of value added per worker, own €40,000 in fixed assets per worker, and hold approximately €20,000 in cash per worker.

Panel B of Table 2.2 compares the characteristics of firms that enter a new sector by acquisition with those that build on their preexisting internal human capital. The results show that firms that build have significantly smaller sales in the sector of entry than firms that buy, with €9.39 million less in sales on average in the year of entry. Building firms are also 13% less likely to stay in the sector of entry at a one-year horizon. They invest €9.95 million less in the year of entry and employ approximately 1,770 fewer workers on average. In addition, firms that diversify by acquisition are more profitable, more capital intensive and have higher internal funds. These significant differences in observable characteristics between firms that build and firms that buy emphasize the importance of including control variables in the empirical analysis.

[Insert [statdesc]Table 2.2 here]

## 2.5. Human capital and corporate diversification

### 2.5.1. Main results

Table 2.3 presents our main results. We test our model's prediction that firms are more likely to build when their human capital is more adapted to operate in the sector of entry (prediction 2.2.2). The dependent variable  $1(Buy)_{g,n,t}$  is a dummy variable equal to one if firm  $f$  enters a new sector  $n$  at time  $t$  through the acquisition of an existing firm and zero if it enters by building on its own resources. The main independent variable *Internal Human Capital* measures the extent to which a firm already employs workers from key occupations for the sector of entry prior to diversification. All our regressions include interacted sector of origin  $\times$  sector of entry  $\times$  year fixed effects. This specification neutralizes potential unobservable time-varying synergies between the sector of entry and that of origin. The idea is to compare firms that operate in the same sector of origin and enter the same new sector in the same year to isolate the effect of human capital on firms' decision to enter through an acquisition or by building on their own resources. All specifications control for firms' size (log number of workers), and we also control for firms' total cash holdings, tangible assets, and value added, with the latter variables being scaled by the number of workers.

Consistent with our model's prediction and Figure 2.2, we find that the internal human capital of the firm prior to entry is negatively correlated with the probability to enter by acquisition. A one-standard-deviation increase in internal human capital is associated with a 1 percentage-point decline in the likelihood of entering

by acquisition (columns 1 and 2). This relationship is sizable, equal to 50% of the unconditional probability of buying, and significant at the 1% level. The point estimate is unchanged when we add control variables in column 2, suggesting that the control variables are uncorrelated with our key dependent variable *Internal Human Capital*. We conclude that firms possessing human capital adapted to the sector of entry are more likely to enter by building on their own resources. The estimates in Table 2.3 also show that the likelihood of buying relative to building increases with firm size. By contrast, cash holdings, tangibility and profitability do not appear to be significantly associated with the mode of entering a new sector.

An alternative explanation is that firms anticipate the mode of entry by adjusting the composition of their internal human capital several years before diversifying. If this were the case, human capital would be endogenous to the mode of entry, and our interpretation would be biased. To overcome this problem, we test whether lagged values of internal human capital also predict the mode of entry. We find that firms' decision to build rather than buy is still negatively correlated with internal capital two and three years before entry (columns 3 and 4, respectively). The point estimates on each lagged measure of human capital remain unchanged, suggesting that firms do not significantly modify their workforce composition during the years preceding entry.

Finally, to address the concern that our results could be confounded by firms having developed a certain expertise for a given being type of entry irrespective of the composition of their internal human capital, we adopt a within-firm estimation in column 5. Again comparing firms that operate in the same sector of origin and enter the same new sector in the same year (i.e., with sector of origin  $\times$  sector of entry  $\times$  year fixed effects), we find that within firms, the magnitude of the relationship between internal human capital and the likelihood of entering by acquisition remains significant at the 5% level. However, the economic magnitude of this relationship is only a third of that across firms within sector specification (column 1). Note that in the within firms specification, the coefficient is identified on firms that perform several diversified entries during the sample period and switch their mode of entry. In addition, the coefficients of firm size and value added per worker in column 5 have a different sign than in the other columns. Our interpretation is that firms that diversify multiple times tend to buy at a stage of their lifecycle in which they are relatively smaller and less profitable.

To conclude, Table 2.3 supports our model's prediction 2.2.2; that is, firms choose to build rather than buy when their existing workforce is more adapted to operate into the sector of entry. According to our model, one rationale for this result is that hiring workers in key occupations is too costly because of tight LLMs. We discuss the role of labor market tightness in Section 2.8.1.

[Insert [tablemain]Table 2.3 here]

### 2.5.2. The role of size and financial constraints

Firms that buy tend to be larger (Table 2.2). One potential concern is that only large firms may be able to pay the fixed costs associated with an acquisition. Hence, small firms would always end up entering by building, and human capital considerations would be irrelevant. We should expect in that case no significant relationship between human capital and the type of entry for smaller firms. In columns 1 and 2 of Table 2.4, we interact our human capital measure with firm size. We use terciles of firm size (number of workers) to allow for non-linear effects of firm size on the decision to build or buy. We find that the interaction terms are not significantly different from zero, suggesting that human capital plays a role in the decision to build or buy for both small and large firms.

A potential explanation for why larger firms tend to grow more by acquisition is that they are less financially constrained than smaller firms. In columns 3 and 4, we test whether our measure of internal human capital interacts with internal financial resources as proxied by cash holdings. The results show that the relationship between internal human capital and the probability of building or buying does not differ across firms with different levels of cash earnings.

Finally, public firms may have different growth strategies than private firms because they can raise equity to fund acquisitions. We test whether the relationship between internal human capital and the decision to build or buy differs for firms that are privately and publicly owned. We re-estimate our baseline specification separately on public and private firms in columns 5 and 6 of Table 2.4.<sup>35</sup> We find that the negative relationship holds both for public and private firms. Note that given the small number of public firms in France, we only perform those regressions with sector of origin  $\times$  sector of entry interacted separately with fixed effects and year fixed effects. In summary, Table 2.4 confirms that the main findings hold across different types of firms.

[Insert [tablecashsizeownership]Table 2.4 here]

---

<sup>35</sup>In the Appendix, Table 2.A7 interacts our covariates with a dummy equal to 1 if the target is publicly listed. Our main results on the role of internal human capital are unchanged, although we find public targets to be acquired relatively more when their ratio of fixed assets over total number of workers is lower.

### 2.5.3. The role of sectoral and geographical distances

Firms do not diversify in random sectors but consider complementarities and synergies between the sectors of origin and entry.<sup>36</sup> Specifically, one may argue that firms choose to diversify by acquisition in more distant sectors from the sector of origin and may instead enter close sectors by building, irrespective of labor considerations. Geographical distance between the firm and the market of entry can also play an important role, with physically more distant firms lacking the local resources to enter a market by building on their existing resources. Instead, we expect firms are more likely to build with their existing workforce if they are physically close, because it is less costly for them to get their current employees to travel and/or relocate.

First, we investigate the effect of product market distance on a firm's decision to build or buy. We interact internal human capital with a firm-level measure of the firm's distance to the sector of entry. This distance is based on the correlation between the product market portfolio of a firm prior to entry and product market portfolios of firms already operating in a sector (see Section 2.4.2 for details). A distance close to zero, for instance, means that the entering firm has very similar activities to incumbent firms.

In column 1, we use terciles of the distance, and in column 2, we interact them with the measure of internal human capital in column 2 to investigate how human capital interacts with the product market distance between the sector of origin and the sector of entry. We find that firms tend to buy more frequently in more distant product markets (third tercile). The point estimate for human capital is very close to that in Table 2.3, which suggests that workforce composition is not explained by sectoral similarities. Interestingly, in column 2, we find that the internal human capital coefficient is stronger for firms in the 2<sup>nd</sup> and 3<sup>rd</sup> terciles of distance. This finding suggests that diversification in distant product markets amplifies the importance of labor costs in the decision to build or buy.

Second, in columns 3 and 4 of Table 2.5, we test the effect of a vertical integration in the decision to build or buy. Labor cost considerations may be less relevant in the presence of important vertical links. Firms might be willing to enter upstream sectors to acquire suppliers and to facilitate transfers of goods along the production chain. To account for the firm's position in the production chain, we measure vertical integration following Fan and Goyal (2006). Using the 1995 input-output (IO) matrix for France compiled by the Organization for Economic Cooperation and Development (OECD), we identify vertical links between a firm and the sector of

---

<sup>36</sup>Hoberg and Phillips (2010) show that product market distance plays an important role in determining M&A patterns. Firms that make similar products tend to merge more with one another, enhancing more value creation. The authors develop a measure based on textual analysis of 10k filings available for publicly listed firms in the US.

entry when more than 5% of the inputs used by the sector of origin come from the sector of entry (we also use 1%, 10% and 20% as thresholds).<sup>37</sup> Because the variable is defined at the level of a sector of origin and the sector of entry, we replace the interacted fixed effects with separate origin, entry and year fixed effects. We do not find any systematic relationship between the presence of vertical links and the decision to build or buy. In particular, firms that start operations in an upstream sector do not seem to enter more often by acquisition.

Third, to investigate the role of the geographical distance between the firm and the market of entry, we construct a dummy variable, *New geographic zone*, that is equal to 1 if the firm diversifies in a new geographic area.<sup>38</sup> 4.2% of entries by acquisition occur in a new geographic area, and only 1.8% of entries by building. In column 5 of Table 2.5, we find that an entry in a new geographic area is positively associated with the likelihood to diversify by acquisition. Hence, physically distant firms are more likely to buy whereas firms that stay in the same area are more likely to build on their existing workforce. Besides, we do not find the economic or statistical significance of our internal human capital measure to be affected when we control for the entry in a new geographic area. In column 6, we do not find that the interaction of physical distance to interact with human capital is statistically different from zero. Overall, the evidence suggests that the role of physical distance in a firm's decision to build or buy is distinct from that of internal human capital.

[Insert [tabledist]Table 2.5 here]

## 2.6. Robustness checks and alternative mechanisms

### 2.6.1. Alternative measures of human capital

In this section, we check the robustness of our results to alternative definitions of the measure of internal human capital. To do so, we replicate the main specification in Table 2.3 using different versions of *Internal human capital*. Table 2.6 reports the results. In column 1, the main independent variable is a dummy variable taking value one if the firm does not employ any worker in the top-10 most important occupations for the sector of entry. The ranking of occupations within sectors is based on the estimated values of the occupation  $\times$  sector  $\times$  fixed effects (see Section 2.3). We find that firms that do not hire any worker in the top-10 occupations for the sector of entry are 40% more likely to buy (column 1).

---

<sup>37</sup>Note that the IO matrix from France that we use is rather coarse (35 industries by 35 industries). The regression excludes observations for which the industry of origin and the industry of entry are the same because the vertical link variable is not defined in that case.

<sup>38</sup>In untabulated results, we find similar results when defining a geographic zone using departments instead of regions.

In column 2 of Table 2.6, we take the weighted average of the fixed effects with weights equal to the share of workers in a given occupation in the workforce of the firm. This alternative measure of internal human capital assumes that the firm allocates workers to the new sector in proportion to the existing occupational structure (see Section 2.3.2). In column 3, we exclude CEOs from the set of occupations used to build the measure of internal human capital. One concern with CEOs' wages is that they may be determined by factors other than their contribution to the firms' performance, for instance, moral hazard, or information asymmetries.

Finally, in column 4, we change the unit of observation used in the estimation of the occupation  $\times$  sector  $\times$  year fixed effects from firms to plants. Plants are supposed to be less diversified entities than firms, thus estimating the fixed effects at the plant level should yield more precise estimates. Indeed, when we estimate fixed effects at the firm level, we make the implicit assumption that every worker is involved in the production process of the firm's main activity.

Our main result is not affected by any of these alternative measures of internal human capital: The negative relationship between human capital and the probability to buy remains economically and statistically significant. Moreover, the point estimates are very similar across specifications.

[Insert [tablerob]Table 2.6 here]

### 2.6.2. Does selection into diversification drive our results?

A potential limitation of our analysis is that we focus on the build versus buy trade-off without eliciting the decision to diversify in the first place. If the entry and the type of entry in the new sector are jointly determined and driven by unobservable factors, the OLS estimates may be biased.

Our approach to address this issue is to identify plausible scenarios in which self-selection could invalidate our results.<sup>39</sup> First, firms with under-performing business segments may be willing to shift their activities to sectors with better prospects. Thus, firms are likely to choose sectors in which they can easily redeploy their existing workforce. In this scenario, the likelihood of an entry by acquisition would also be negatively related to internal human capital.

---

<sup>39</sup>A classic solution to address selection issues in diversifying choices is to find an instrumental variable that affects the choice of sector but is plausibly orthogonal to the main dependent variable. For instance, Tate and Yang (2016b) uses Tobin's Q as an instrument in a two-stage Heckman selection model. Applied to our context, we would need an instrument that affects the choice to diversify but not the type of entry. In addition, this instrument would need to be sector-specific because we would have to instrument not only for the choice to build or buy, but also for the decision to diversify in a given sector. Finding such an instrument appears to be a difficult task; therefore, we choose to focus on specific scenarios in which such selection issues are likely to arise.

We test the influence of sectoral reallocation in Table 2.7. We identify “shifting firms” as those firms that enter a new sector while using internal resources to shift a substantial part of their activities. Under the tested hypothesis, the negative coefficient for human capital on the likelihood to enter by building would be driven by the sub-sample of shifting firms. Excluding them should result in a non-significant coefficient for human capital. In contrast, the results in columns (1) to (3) show that the point estimates for our main variable remain unchanged.

Consider then the other polar case. Firms with a specific expertise or organizational capital for acquisition may always be willing to always diversify by acquisition irrespective of internal human capital considerations. Since firms are unlikely to employ the right set of workers to enter a new sector if they never expect to build using their existing resources, this could translate into an observed negative relationship between internal human capital and the propensity to buy. To address this particular issue, in column (4) of Table 2.7, we focus on firms that enter multiple sectors by acquisition (“serial acquirers”).<sup>40</sup> We still find a significant, negative coefficient for human capital when excluding those serial acquirers. Overall, these findings about serial acquirers and firms likely to shift their activities to sectors with better prospects suggest a limited role for selection issues.

[Insert [tableserial]Table 2.7 here]

### 2.6.3. The role of scale and physical capital

We investigate the issue of the scale of the new activity in the sector of entry. If the entry is small, existing workforce slack can be used for it conditional on the workforce having the right skills. It is also the case that if the entry is small, the firm does not have to hire many workers. When size is large, it becomes much trickier to enter the new sector by building on existing resources, because it requires workers who can function together, so hiring one worker at a time can be inefficient and time-consuming. Another related issue is the role of physical capital adjustment (e.g., equipment, or machinery), which may act as a confounding factor in the relationship between human capital and firms’ decision to build or buy.

First, to investigate the role of the scale of the new activity, we compare the build or buy decision of firms that realized similar entry sales in the year of entry. As firms should enter a new sector only if they anticipate high enough entry sales to offset entry costs, we use entry sales to proxy for entry costs. In columns 1 and 2 of Table 2.8, we rank entry sales into 10 deciles and run our baseline regressions with interacted sector of origin  $\times$  sector of entry  $\times$  year  $\times$  sales decile fixed effects. This

---

<sup>40</sup>Among firms that enter a sector using external resources at least once, 50% are classified as serial acquirers.

specification allows us to compare firms that operate in the same sector of origin, enter the same sector in the same year, and make similar sales when entering the new sector. Although this specification creates many singletons that are dropped from the sample (the number of observations drops from 75k to 45k), our point estimates remain nearly unchanged.

Second and similarly, we proxy for physical capital adjustments using the volume of capital expenditures made by firms when entering the new sector. To do so, we compare firms investing similar amounts in the year of entry. For firms that build, we measure investment using capital expenditures in the year of entry. For firms that buy, we measure investment as the amount of fixed assets in the target. We run our baseline regression with the interacted sector of origin  $\times$  sector of entry  $\times$  year  $\times$  investment decile fixed effects. The idea is to isolate the effect of human capital on the build or buy decision, irrespective of differences in capital expenditures. In columns 3 and 4 of Table 2.8, we find that the economic magnitude of the role of human capital is reduced by approximately half. It remains negative and statistically significant at the 5% level, with or without the inclusion of control variables. Based on the results in Table 2.8, we conclude that human capital matters in firms' decision to build or buy, irrespective of the scale of the new activity.

[Insert [tablelaborcapital]Table 2.8 here]

## 2.7. Evidence of labor adjustment costs

The previous sections established that firms that do not employ the right set of workers tend to enter a new sector by acquisition. This finding, we argue, suggests that firms prefer to pay the costs associated with acquiring and restructuring a target when the costs of adjusting the existing workforce are large. In this section, we focus on firms that diversify by building on their existing resources to highlight the existence of such adjustment costs.

### 2.7.1. Within-firm human capital and diversification choice

We define a firm as a set of subsidiaries. Therefore, the composition of the workforce may vary within a firm, with some subsidiaries being better prepared to enter the new sector than others. Within firms that build, we should expect the entry to be made through subsidiaries that already employ the right set of occupations to minimize reallocation costs in the internal labor market. We construct our measure of internal human capital for each subsidiary and estimate the following model:

$$\mathbb{1}(\text{Build})_{f,n,t} = \lambda_{g,n,t} + \beta \cdot \text{Internal Human Capital}_{f,n,t-1} + \gamma \cdot X_{f,n,t-1} + \epsilon_{f,n,t}$$

The dependent variable is a dummy variable that takes value one if the entry into the new sector is made through subsidiary  $f$ , zero otherwise. We include firm  $\times$  sector of entry  $\times$  year fixed effects to compare the internal capital of the different subsidiaries of the same firm  $g$ . This specification leads mechanically to the exclusion of stand-alone firms because we focus on the heterogeneity of human capital across subsidiaries.

In column 1 of Table 2.9, we find that within a firm, entry is more likely to be achieved through a subsidiary with the appropriate set of occupations for the new sector. In column 2, we add control variables at the subsidiary level. The estimates show that large, productive and cash-rich subsidiaries are more likely to diversify in a new sector. In addition, in columns 3 and 4, we show that lagged values of internal human capital are also positively correlated with the entry dummy. As for firms, internal human capital at the subsidiary level is sticky over time, suggesting that the composition of the workforce is not adjusted in anticipation of diversification. Overall, these findings are consistent with the presence of reallocation costs in the internal labor market.

[Insert [tableinternal]Table 2.9 here]

### **2.7.2. Internal human capital and workforce adjustment**

Our measure of human capital captures the extent to which the firm's workforce is adapted to the sector of entry. Therefore, among firms that enter a new sector by developing their own resources, we should find that firms with lower human capital hire relatively more workers to adjust their workforce. We test this hypothesis by examining employment growth within subsidiaries. Precisely, we focus on the subsidiaries that, within diversifying firms, begin selling in the new sector. This choice allows us to observe more precisely the adjustment in labor associated with entry because it allows us to abstract from employment variations in the other subsidiaries.

In Table 2.10, we demonstrate the existence of a negative relationship between internal human capital at  $t - 1$  and the growth rate of employment between  $t - 1$  and  $t + 1$  (with  $t$  being the year of entry into the new sector). In column 1, we find that the higher the internal human capital is, the less additional workers the subsidiary hires to operate in the new sector. In columns 2 and 3, we examine the timing of new hiring. We find that subsidiaries with lower internal human capital do not hire more workers prior to entry (column 2). Instead, subsidiaries tend to adjust their workforce after having entered the new sector (column 3). In column 4, we show that the newly hired workers are relatively more adapted to the sector of entry, because they imply an increase in firms' human capital. Our interpretation

is that when entering a new sector by developing their own resources, firms adjust their internal resources for the new sector.

[Insert [tablerealloc]Table 2.10 here]

## 2.8. The role of labor market frictions

### 2.8.1. The effects of local labor market tightness for key occupations

Given the costs associated with an acquisition, why do firms not hire new workers instead of buying an existing firm? The model predicts that firms are more likely to buy than to build on their existing workforce when workers in key occupations are in short supply in the external labor market (prediction 2.2.2). Specifically, the negative relationship between internal human capital and the likelihood of buying should be stronger when it is costly for firms to hire workers in the external labor market. In this section, we empirically assess the role of LLM tightness in the decision to build or buy using the *Local Labor Tightness* variable described in 2.4.2.

Columns 1 and 2 of Table 2.11 show that the point estimates on the second and third terciles of LLM tightness are positively related to the decision to buy. Thus, firms are significantly more likely to buy when labor markets for key occupations are tighter. Moreover, the interaction terms show that the link between internal capital and the type of entry is stronger in the presence of greater market frictions.

These findings hold both with and without control variables (columns 1 and 2). We confirm this finding when we divide the sample by tercile of LLM tightness and re-estimate the regression separately on each sub-sample: The coefficient on internal human capital increases in absolute value and is significant at 10% only for the last tercile of LLM tightness. Overall, the finding is consistent with the prediction that human capital determines firms' decision to build or buy and is especially acute in the presence of labor market frictions.

[Insert [tabletightness]Table 2.11 here]

To check the robustness of this result, we conduct the same analysis using an alternative measure of LLM tightness. We use the definition in Equation (33) in Section 2.4.2, based on the percentage of occupations in short supply. Table 2.A6 presents the results, which are very similar to our earlier findings: Firms buy more when LLMs are tight, and the role of human capital is stronger when LLMs are tight. When using the interacted sector of origin  $\times$  sector of entry  $\times$  (geographical) LLM fixed effects, the results still hold, but only significantly so in the tightest LLMs due to a lack of statistical power.

### 2.8.2. Local labor market tightness and the value of building

According to our model, firms' internal resources determine expected profits from entering a new sector by building versus buying. Firms choose to build when their workforce is adapted to the sector of entry, which is more profitable than buying an existing company. Therefore, our next question is whether firms that enter by building with higher internal human capital create value, as implied by our model. Are firms that do not and face greater hiring costs less productive in the short run? Based on predictions 2.2.2 and 2.2.2, we should expect that the profits firms can generate from building are positively related to internal human capital. Moreover, this relationship should be stronger when key occupations are in short supply in the external labor market. In this last section, we directly examine the relationship between the value of building and the human capital of the firm.

We proxy the value of building by studying (the logarithm of) entry sales after the entry. In columns 1 and 2 of Table 2.12, we find that entry sales are larger when the workforce of the firm is more adapted to the sector of entry. This finding suggests that the value of building is higher when the firm has adequate internal resources. In columns 3 to 5, we run the analysis on a subsample of LLM tightness terciles and show that the positive link is entirely driven by tight labor markets (3rd tercile, column 5). This finding suggests that the value of building depends on firms' internal human capital when LLM frictions are important. In other words, firms' internal resources matter only when it is costly to obtain them outside the firm.

[Insert [tablevalue]Table 2.12 here]

### 2.8.3. Build or buy and firing costs

We test whether firing costs affects firms' decision to build or buy. According to our model, when a firm buys to enter a new sector, it has to incur the cost of restructuring the target (see Appendix 2.11.1). Therefore, our first hypothesis is that the higher the restructuring cost, the less attractive is the option to buy. Furthermore, the more adapted an acquiring firm's internal human capital, the more workers must be laid off after the acquisition because of a higher overlap of key worker occupations; therefore, the higher will be the restructuring cost after the acquisition. Therefore, our second hypothesis is that a higher firing cost makes the option to buy less attractive for firms with adapted internal human capital.

We perform two tests of these hypotheses. First, we use the fraction of permanent to temporary workers in a firm's workforce to proxy for firing costs, hence the cost at which firms can restructure their workforce. The higher the fraction of permanent workers in a firm's workforce, the less maneuver a firm has to replace temporary workers with workers that are more adapted for the sector of entry.

Besides, in France, laying-off a permanent worker involves higher reparation costs in case of prejudice (Fraisie, Kramarz and Prost, 2015). Therefore, expected firing costs increase in the fraction of permanent workers in the workforce.

In columns 1 and 2 of Table 2.13, we test whether the fraction of permanent employees in a firm's workforce is correlated with firms' decision to build or buy. We find support for our first hypothesis: In column 1, we find a positive relationship between the fraction of permanent workers in a firm's workforce and the decision to build, though this relationship is not significant. However, in line with our second hypothesis, this fraction interacts significantly with our measure of internal human capital (column 2), implying a higher firing cost increases the importance of internal human capital in firms' decision to build or buy.

One potential problem with this first proxy of firing costs is that firms decide whether they want to hire workers under permanent or temporary contracts. Therefore, in a second set of tests, we use the geographic variation in the average length of local labor cases settlements across jurisdictions as a proxy for firing costs that is exogenous to firms (see Sections 2.4.1 and 2.4.2). In column 3 of Table 2.13, we find support for our first hypothesis: Firms are less likely to enter a new market by acquisition in areas where the average length of labor case settlements is high (though this relationship is not statistically significant). In line with our second hypothesis, we find that internal human capital matters significantly more in firms' decision to build or buy when the firing cost is high (column 4). Overall, the results in Table 2.13 show that firing costs amplify the importance of internal human capital in firms' decision to build or buy.

[Insert [tablefiringcosts]Table 2.13 here]

## 2.9. Conclusion

Why do some firms enter a new sector by acquiring an existing company ("buy"), while others do so using their existing resources ("build")? When a firm buys to enter a new sector, it has to incur both the costs of acquiring and restructuring the target, but also secures access to the target's productive resources. When a firm builds on its existing resources to enter a new sector, it must pay the adjustment costs needed to acquire an adapted set of capabilities.

We focus on the role of labor and construct a firm-level measure of human capital based on the occupational structure of the workforce. Our main explanatory variable measures the extent to which the firm's internal human capital is adapted to the sector of entry.

We show that the vast majority of entries in a new sector consist of firms that

build on their internal resources, and that firms choose to buy an existing company when their human capital is not adapted to the sector of entry.

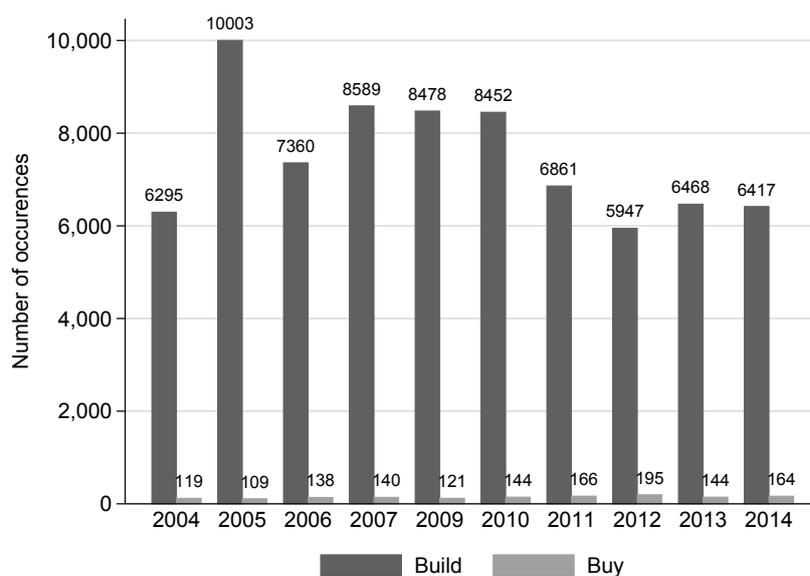
We find evidence that labor adjustment costs contribute to the importance of internal human capital in firms' choice to build or buy. On the one hand, firms are more likely to buy when it is costly to hire key workers on the external job market, i.e., when key worker occupations are in short supply. On the other hand, higher firing costs make the option to buy less attractive for firms with adapted internal human capital.

Our findings imply fundamental factors are driving firms' decision to build or buy. They are consistent with firms choosing organizational structures that best deploy the economy's pool of specialized resources (Maksimovic and Phillips, 2002). Thereby, they contribute to the literature on corporate diversification by showing that both the set of internal resources and the cost of accessing external resources play a role in explaining how firms diversify.

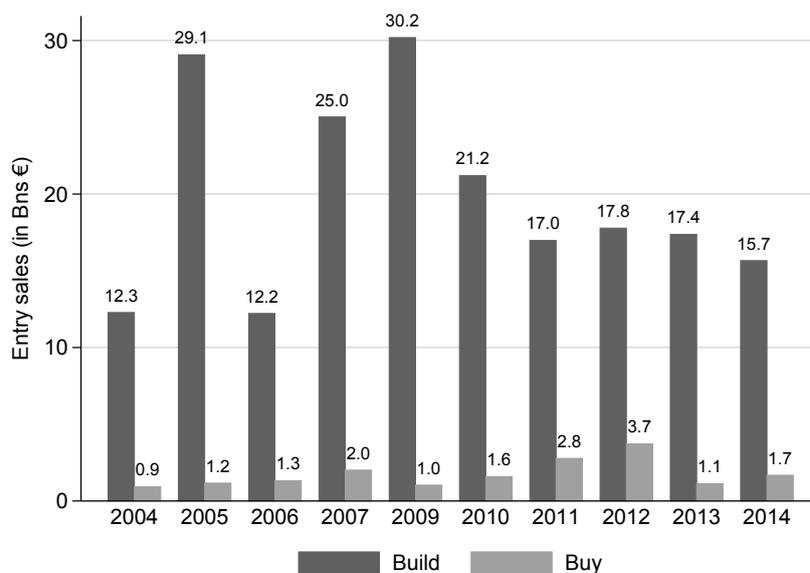
## **2.10. Figures and Tables**

**Figure 2.1. Number and Size of Build and Buy Entries**

*Source:* SDC Platinum, BvD Zephyr, ESA survey. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. This figure displays the number of external (Buy) and internal (Build) entries by year and the aggregate sales by type of entry. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Entries are identified with sales reported at the 5-digit level of the French SIC.



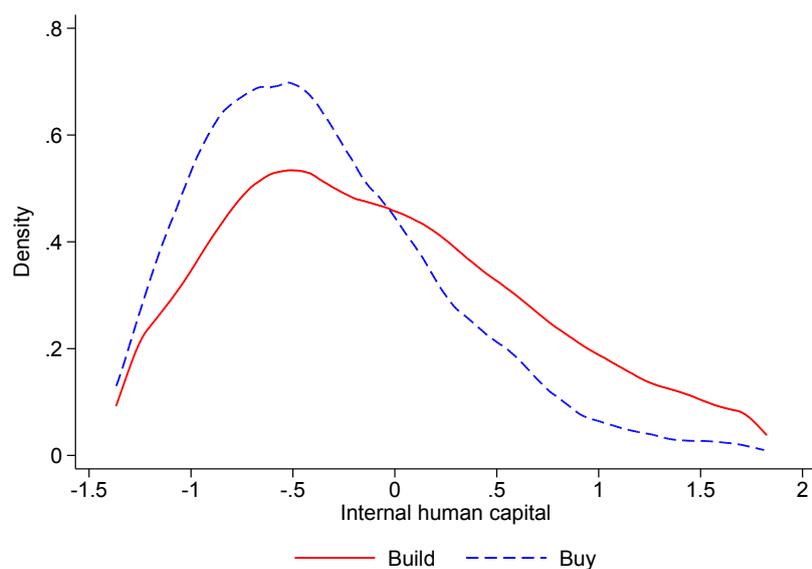
(a) Number of entries



(b) Aggregate sales by type of entry

**Figure 2.2. Human Capital by Type of Entry**

The figure displays the probability distribution function of *Internal Human capital* by mode of entry. *Source*: SDC Platinum, BvD Zephyr, EAE survey, matched employer-employee dataset. *Sample*: Firms that enter a new sector either internally (build) or externally (buy) during the periods 2003-2007 and 2009-2014. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Build entries are identified using reported sales from the ESA survey at the 5-digits level of the French SIC. *Internal Human Capital* $_{g,n,t-1}$  measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2). *Internal Human Capital* $_{g,n,t-1}$  is trimmed at the 5% level (for this graph only).


**Table 2.1. Evolution of the Numbers of Build and Buy Entries**

*Source*: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset. *Sample*: Firms that enter a new sector during the periods 2003-2007 and 2009-2014. This table reports the ratio of Buy entries to Build entries. A firm is said to “buy” when it enters a new sector through a M&A (source: SDC Platinum and Bureau van Dijk Zephyr). A firm is said to “build” when it enters a new sector through one its existing subsidiaries (source: ESA survey).

Panel A. Buy and Build entries by year

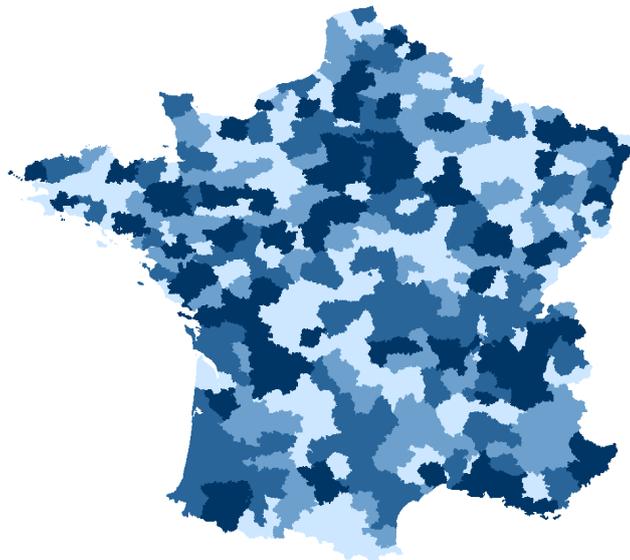
	2004	2005	2006	2007	2009	2010	2011	2012	2013	2014
Build (number)	6295	10003	7360	8589	8478	8452	6861	5947	6468	6417
Buy (number)	119	109	138	140	121	144	166	195	144	164
Buy (% , frequency)	1.85	1.08	1.84	1.60	1.4	1.67	2.36	3.17	2.18	2.49
Buy (% , sales)	6.98	3.85	9.79	7.44	3.31	6.95	14.07	17.30	6.06	9.68

Panel B. Buy and Build entries for varying definitions of sector

Industry level:	5 digits	4 digits	3 digits	2 digits	1 digit
Build (number)	74,870	72,250	63,176	49,570	32,431
Buy (number)	1,440	1,363	1,163	941	584
Buy (% , frequency)	1.89	1.85	1.81	1.86	1.77
Buy (% , sales)	8.05	8.55	8.31	10.57	15.72

### Figure 2.3. Occupations in Short Supply

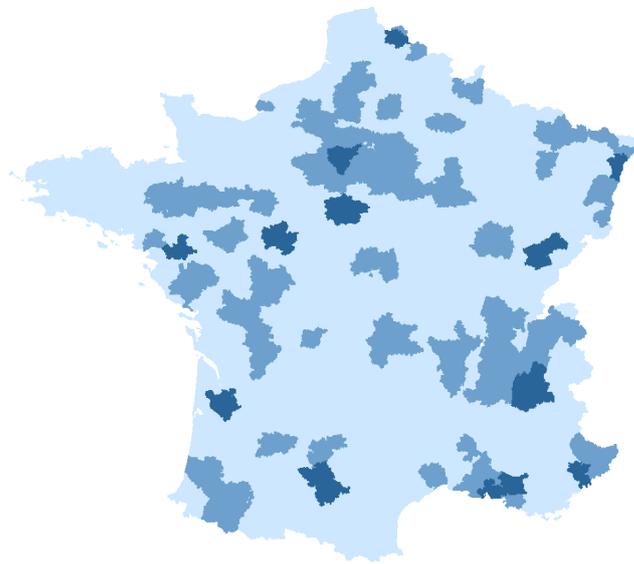
*Source:* French national unemployment agency. Figure 2.3 plots the distribution of the tightness of labor markets in 2013. Labor market tightness is measured by the number of occupations in short supply in a given local labor market. Darker shades of blue indicate a higher degree of tension in the local labor market. Occupations in short supply are identified by the French national employment agency as occupations for which (i) the ratio of job offers over job applications is high (ii) surveyed employers forecast that it will be difficult to fill posted offers. There are 348 different local labor markets.



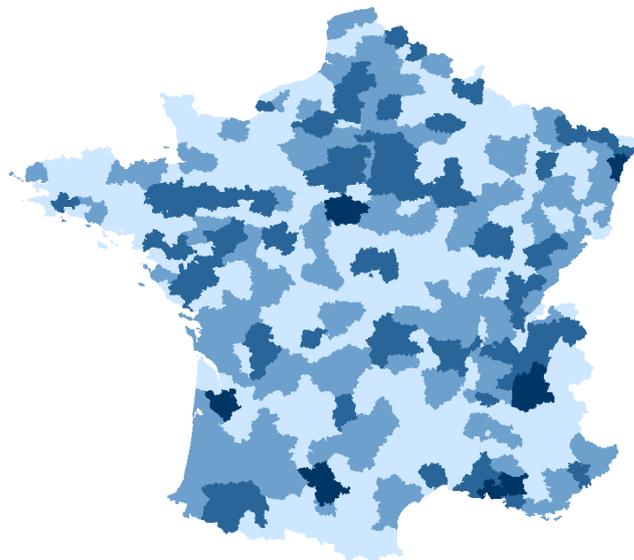
(a) Number of occupations in short supply by local labor market.

**Figure 2.4. Local Labor Markets Tightness**

*Source:* French national unemployment agency. The figures plot the value of  $LLM\ Tightness_{n,z,t}$  of each local labor market  $z$  in  $t = 2013$  for different sectors  $n$ . Figure 2.4a focuses on the manufacture of pharmaceutical preparations and figure 2.4b focuses on the manufacture of motor vehicles. Darker shades of blue indicate local labor markets with a larger number of occupations in short supply. Occupations in short supply are identified by the French national employment agency as occupations for which (i) the ratio of job offers over job applications is high (ii) surveyed employers forecast that it will be difficult to fill posted offers. France is divided into 348 different local labor markets  $z$ .



(a) Manufacture of pharmaceutical preparations



(b) Manufacture of motor vehicles

**Table 2.2. Summary Statistics**

*Source:* SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. This table reports descriptive statistics for firms that are identified to enter a new sector, either internally or externally, during the periods 2003-2007 and 2009-2014. Panel A reports the distribution of the main firm characteristics. Panel B compares the means firms' characteristics that enter a new sector internally and those that enter by acquisition. A firm is said to "buy" when it enters a new sector through a M&A (source: SDC Platinum and Bureau van Dijk Zephyr). A firm is said to "build" when it enters a new sector by one its existing subsidiaries (source: ESA survey). Sectors refer to an industry at the 5-digits level of the French SIC. Description of the variables are reported in Appendix C.

Panel A. Distribution of firms characteristics

	N	Mean	St.Dev.	Percentiles				
				5 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	95 <sup>th</sup>
$\mathbb{1}(\text{Buy})_{g,n,t}$	76310	0.02	0.14	0.00	0.00	0.00	0.00	0.00
Entry sales $_{g,n,t}$ (M euros)	76310	2.82	21.32	0.00	0.05	0.27	1.21	8.95
$\mathbb{1}(\text{1-year survival})_{g,n,t+1}$	76310	0.47	0.50	0.00	0.00	0.00	1.00	1.00
Investment $_{g,n,t}$ (M euros)	76310	0.96	16.24	0.00	0.03	0.10	0.35	2.44
Internal Human Capital $_{g,n,t-1}$	76310	0.14	1.04	-1.29	-0.59	-0.00	0.69	2.05
#workers $_{g,t-1}$ (in thousands)	76310	0.63	5.16	0.02	0.04	0.07	0.19	1.29
Value added/N.workers $_{g,t-1}$	76310	0.05	0.03	0.02	0.03	0.04	0.06	0.11
Fixed assets/N.workers $_{g,t-1}$	76310	0.04	0.05	0.00	0.01	0.03	0.06	0.14
Cash holdings/N.workers $_{g,t-1}$	76310	0.02	0.02	0.00	0.00	0.01	0.02	0.06
Product market distance $_{g,n,t-1}$	76310	0.87	0.15	0.53	0.81	0.94	0.98	1.00
Vertical integration $_{g,n,t-1}$	44947	0.67	0.47	0.00	0.00	1.00	1.00	1.00

Panel B. Comparison of build and buy entries

	Build		Buy		Difference	
	Mean	St.Dev.	Mean	St.Dev.	Mean	p-value
Entry sales $_{g,n,t}$ (M euros)	2.64	20.52	12.04	45.85	-9.39***	(0.00)
$\mathbb{1}(\text{1-year survival})_{g,n,t+1}$	0.47	0.50	0.59	0.49	-0.13***	(0.00)
Investment $_{g,n,t}$ (M euros)	0.78	15.35	10.72	40.43	-9.95***	(0.00)
Internal Human Capital $_{g,n,t-1}$	0.15	1.04	-0.21	0.77	0.36***	(0.00)
N.workers $_{g,t-1}$ (in thousands)	0.59	4.82	2.37	14.33	-1.77***	(0.00)
Value added/N.workers $_{g,t-1}$	0.05	0.03	0.07	0.04	-0.01***	(0.00)
Fixed assets/N.workers $_{g,t-1}$	0.04	0.05	0.05	0.06	-0.00	(0.16)
Cash holdings/N.workers $_{g,t-1}$	0.02	0.02	0.02	0.02	-0.01***	(0.00)
Product market distance $_{g,n,t-1}$	0.87	0.16	0.92	0.10	-0.05***	(0.00)
Vertical link $_{g,n,t-1}$	0.68	0.47	0.59	0.49	0.08***	(0.00)
Observations	74,870		1,440		76,310	

**Table 2.3. Human Capital and Corporate Diversification**

*Source:* SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and analyses the role of human capital on the type of diversification strategy. The dependent variable  $\mathbb{1}(\text{Buy})_{g,n,t}$  is a dummy variable that takes the value one if the entry of firm  $g$  in sector  $n$  at time  $t$  is made through an acquisition, zero if the entry is made internally. Entries are identified with sales reported at the 5-digit level of the French SIC. The main independent variable is a firm-level measure of human capital *Internal Human Capital* $_{g,n,t-1}$ . The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2). Control variables include the number of workers in logarithm as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include Sector Origin  $\times$  Entry  $\times$  Year fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$				
	(1)	(2)	(3)	(4)	(5)
Internal HC $_{g,n,t-1}$	-0.015*** (0.004)	-0.010*** (0.003)			-0.005** (0.002)
$\log(\#\text{workers})_{g,t-1}$		0.010*** (0.002)	0.010*** (0.002)	0.011*** (0.002)	-0.045*** (0.013)
Cash holdings/ $\#\text{workers}_{g,t-1}$		0.044 (0.045)	0.043 (0.050)	0.054 (0.058)	-0.038 (0.145)
Fixed assets/ $\#\text{workers}_{g,t-1}$		-0.021 (0.028)	-0.024 (0.031)	-0.023 (0.034)	-0.180* (0.106)
Value added/ $\#\text{workers}_{g,t-1}$		0.049 (0.035)	0.046 (0.038)	0.038 (0.043)	-0.295*** (0.090)
Internal HC $_{g,n,t-2}$			-0.011*** (0.003)		
Internal HC $_{g,n,t-3}$				-0.014*** (0.004)	
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No	Yes
$R^2$	0.199	0.206	0.208	0.207	0.564
Observations	76354	76296	66145	54230	57923

**Table 2.4. The Role of Size and Financial Constraints**

*Source:* SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the effect of firms' financial constraints on the relationship between human capital and the type of diversification strategy. Three proxy of financial constraints are considered: size, cash holding and public ownership status. Columns (1) and (2) include the second and third terciles of size (number of workers) interacted with *Internal Human Capital* $_{g,n,t-1}$ . Columns (3) and (4) include the second and third terciles of cash holdings over workers (*Cash holdings/N. workers*) interacted with *Internal Human Capital* $_{g,n,t-1}$ . In columns (5) and (6), we split firms between publicly and privately owned firms. Public firms are those that include at least one publicly listed subsidiary within the firm. The dependent variable  $\mathbb{1}(Buy)_{g,n,t}$  is a dummy variable that takes the value one if the entry of firm  $g$  in sector  $n$  at time  $t$  is made through an acquisition, zero if the entry is made internally. Entries are identified at the 5-digit level of the French SIC. The main independent variable is a firm-level measure of human capital *Internal Human Capital* $_{g,n,t-1}$ . The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2). Control variables include the number of workers in logarithm as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(Buy)_{g,n,t}$					
	Size		Cash/N. workers		Type of firm	
	(1)	(2)	(3)	(4)	Public (5)	Private (6)
Internal HC $_{g,n,t-1}$	-0.011*** (0.003)	-0.011*** (0.003)	-0.014*** (0.003)	-0.010*** (0.003)	-0.123** (0.051)	-0.007*** (0.002)
2nd tercile of #workers $_{g,t-1}$	0.005** (0.002)	0.005** (0.002)				
3rd tercile of #workers $_{g,t-1}$	0.025*** (0.005)	0.025*** (0.006)				
2nd t. # workers $_{g,t-1} \times$ Int. HC $_{g,n,t-1}$	-0.001 (0.001)	-0.001 (0.001)				
3rd t. # workers $_{g,t-1} \times$ Int. HC $_{g,n,t-1}$	-0.004 (0.003)	-0.004 (0.003)				
2nd tercile of Cash $_{g,t-1}$			0.002 (0.002)	0.001 (0.001)		
3rd tercile of Cash $_{g,t-1}$			0.004** (0.002)	0.003 (0.002)		
2nd t. Cash $_{g,t-1} \times$ Int. HC $_{g,n,t-1}$			-0.000 (0.001)	0.000 (0.001)		
3rd t. Cash $_{g,t-1} \times$ Int. HC $_{g,n,t-1}$			-0.003* (0.002)	-0.002 (0.002)		
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	No	No
Origin-Entry FE	No	No	No	No	Yes	Yes
Year FE	No	No	No	No	Yes	Yes
Controls	No	Yes	No	No	Yes	Yes
$R^2$	0.204	0.205	0.199	0.206	0.107	0.123
Observations	76354	76296	76354	76354	1198	74204

## 2.11. Appendix

**Table 2.5. The Role of Sectoral and Geographical Distances**

*Source:* SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the effect of complementarities between the sector of origin and sector of entry on the relationship between human capital and the type of diversification strategy. The dependent variable is a dummy variable that takes the value one if the entry of firm  $g$  in sector  $n$  at time  $t$  is made through an acquisition, zero if the entry is made internally. The main independent variable is a firm-level measure of human capital  $Internal\ Human\ Capital_{g,n,t-1}$ . The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2). In addition, columns (1) and (2) include a product market distance adapted from Bloom, Schankerman and Van Reenen (2013). It measures the distance between the sectors in which firm  $g$  operates at  $t-1$  and the sector of entry  $n$ . The distance ranges from 0 to 1 (1 being the maximum). Columns (3) and (4) include the Fan and Goyal (2006)'s measure of vertical relatedness. It measures the intensity of vertical links between the main sector of activity of firm  $g$  at time  $t-1$  and the sector in which  $g$  enters at time  $t$ . "Vertical" is a dummy variable that takes the value 1 if the vertical relatedness exceed 5%. Columns (5) and (6) include a dummy that indicates whether the firm enters a new geographical zone at time  $t$ . We use the "region" as definition of the geographical zone (France is divided in 25 regions over the sample period). A firm enters a new geographical zone if either (i) the entering subsidiary is created at time  $t$  and located in a new region, (ii) the entering subsidiary opens a plant in a new region at time  $t$  or (iii) the entering subsidiary is acquired at time  $t$  and operates in a region in which the firm was not present at  $t-1$ . Control variables include the number of workers in logarithm as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$					
	Product market distance		Vertical integration		Geographical distance	
	(1)	(2)	(3)	(4)	(5)	(6)
Internal HC $_{g,n,t-1}$	-0.010*** (0.003)	-0.006*** (0.002)	-0.019*** (0.005)	-0.018*** (0.005)	-0.015*** (0.004)	-0.010*** (0.003)
2nd tercile of Distance $_{g,n,t-1}$	0.003 (0.003)	0.007** (0.003)				
3rd tercile of Distance $_{g,n,t-1}$	0.009** (0.004)	0.012*** (0.004)				
2rd t. Distance $_{g,n,t-1}$ *Int. HC $_{g,n,t-1}$		-0.007*** (0.002)				
3rd t. Distance $_{g,n,t-1}$ *Int. HC $_{g,n,t-1}$		-0.008*** (0.002)				
Vertical $_{g,n,t-1}$			0.001 (0.006)	0.000 (0.006)		
Vertical $_{g,n,t-1}$ *Int. HC $_{g,n,t-1}$				-0.002 (0.005)		
New geographic zone $_{g,n,t}$					0.013* (0.007)	0.010 (0.008)
New geo. zone $_{g,n,t}$ *Int. HC $_{g,n,t-1}$						-0.006 (0.007)
Origin-Entry-Year FE	Yes	Yes	No	No	Yes	Yes
Origin FE	No	No	Yes	Yes	No	No
Entry FE	No	No	Yes	Yes	No	No
Year FE	No	No	Yes	Yes	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.208	0.208	0.102	0.102	0.201	0.208
Observations	76226	76226	68531	68531	76223	76223

**Table 2.6. Alternative Measures of Human Capital**

*Source:* SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and assesses how our results are robust to the definition of human capital. The dependent variable  $\mathbb{1}(\text{Buy})_{g,n,t}$  is a dummy variable that takes the value one if the entry of firm  $g$  in sector  $n$  at time  $t$  is made through an acquisition, zero if the entry is made internally. Entries are identified at the 5-digit level of the French SIC. We test several alternative definition of our main independent variable  $\text{Internal Human Capital}_{g,n,t-1}$ . In the baseline definition, the measure is computed as the sum of occupation  $\times$  sector  $\times$  year fixed effects present in the workforce of firm  $g$  at time  $t - 1$ . The measure is scaled by the number of occupations in firm  $g$ . The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2). In column (1), the dependent variable is a dummy equal to one if for a sector of entry  $n$  and a time of entry  $t$ , firm  $g$  has no occupation in the top 10 of occupation-sector-year fixed effects. In column (2), the sum of occupation  $\times$  sector  $\times$  year fixed effects is weighted by the number of employees in each occupation in the firm. In column (3), CEO occupations are excluded from the sum of occupations. In column (4), occupation  $\times$  sector  $\times$  year fixed effects are estimated at the plant-level instead of the firm level in the baseline model. Control variables include the number of workers in logarithm as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include sector of origin  $\times$  entry  $\times$  year fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$			
	(1)	(2)	(3)	(4)
$\mathbb{1}(\text{No Top 10})_{g,n,t-1}$	0.008*** (0.003)			
Internal HC $_{g,n,t-1}$ (weighted)		-0.009*** (0.002)		
Internal HC $_{g,n,t-1}$ (no CEO)			-0.010*** (0.003)	
Internal HC $_{g,n,t-1}$ (plant-level)				-0.009*** (0.003)
Origin-Entry-Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
$R^2$	0.205	0.206	0.205	0.206
Observations	76296	75614	76270	76094

**Table 2.7. Selection into Diversification**

*Source:* SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests two potential self-selection scenarios that could drive the main results. The dependent variable  $\mathbb{1}(\text{Buy})_{g,n,t}$  is a dummy variable that takes the value one if the entry of firm  $g$  in sector  $n$  at time  $t$  is made through an acquisition, zero if the entry is made internally. The main independent variable is a firm-level measure of human capital  $\text{Internal Human Capital}_{g,n,t-1}$ . The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2). In columns (1) to (3), firms that decrease their activity in a preexisting sector while entering sector  $n$  (*Shifting firms*) are excluded. *Shifting firms* are firms for which the minimum growth rate of sales is negative and greater than 100%, 50% and 25% in absolute value. We compute the growth rate of sales between  $t-1$  and  $t$  in each sector in which firms were operating at  $t-1$  and take the firm-level minimum of sectoral growth rates. In column (4), *Serial acquirers* are excluded. Serial acquirers are firms that enter more than one sector by acquisition during the time period. Control variables include the number of workers in logarithm as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include sector of origin  $\times$  entry  $\times$  year fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable: Excluding:	$\mathbb{1}(\text{Buy})_{g,n,t}$			
		Shifting firms		
	100%	50%	25%	
	(1)	(2)	(3)	(4)
Internal HC $_{g,n,t-1}$	-0.010*** (0.003)	-0.012*** (0.003)	-0.012*** (0.004)	-0.003*** (0.001)
$\log(\#\text{workers})_{g,t-1}$	0.010*** (0.002)	0.010*** (0.002)	0.011*** (0.003)	0.003*** (0.001)
Cash holdings/ $\#\text{workers}_{g,t-1}$	0.056 (0.046)	0.072 (0.058)	0.073 (0.057)	0.001 (0.014)
Fixed assets/ $\#\text{workers}_{g,t-1}$	-0.010 (0.029)	0.007 (0.032)	0.041 (0.036)	-0.011 (0.010)
Value added/ $\#\text{workers}_{g,t-1}$	0.043 (0.033)	0.039 (0.043)	0.024 (0.050)	0.016 (0.016)
Origin-Entry-Year FE	Yes	Yes	Yes	Yes
$R^2$	0.205	0.222	0.243	0.157
Observations	74510	54072	38847	72681

**Table 2.8. Human Capital and the Scale of the New Entry**

*Source:* SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and give the results of the baseline regression when imposing that firms realize similar entry sales or capital expenditures during the year of entry. In columns (1) and (2), we rank entry sales into ten deciles and run our baseline regressions with interacted sector of origin  $\times$  sector of entry  $\times$  year  $\times$  sales decile fixed effects. Sales are defined as the total amount of sales realized in the sector of entry  $n$  by firm  $g$  at time  $t$ . In columns (3) and (4), we consider replace sales with investment deciles. In the case of build entries, investment is measured as the total amount of capital expenditures realized by firm's firms that entered in sector  $n$  at time  $t$ . In the case of Buy entries, investment is the amount of acquired fixed physical assets. The dependent variable is a dummy variable that takes the value one if the entry of firm  $g$  in sector  $n$  at time  $t$  is made through an acquisition, zero if the entry is made internally. The main independent variable is a firm-level measure of human capital  $Internal\ Human\ Capital_{g,n,t-1}$ . The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2). Control variables include the number of workers in logarithm as well total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$			
	Sales		Investment	
	(1)	(2)	(3)	(4)
Internal $HC_{g,n,t-1}$	-0.009*** (0.003)	-0.007** (0.003)	-0.004* (0.002)	-0.003* (0.002)
Origin-Entry-Year $\times$ Sales bucket FE	Yes	Yes	No	No
Origin-Entry-Year $\times$ Inv. bucket FE	No	No	Yes	Yes
Controls	No	Yes	No	Yes
$R^2$	0.232	0.235	0.267	0.267
Observations	45959	45959	31735	31735

**Table 2.9. Reallocation Costs in the Internal Labor Market**

*Source:* EAE survey, tax files, ownership links dataset, matched employer-employee dataset.  
*Sample:* Subsidiaries of firms that enter a new sector internally during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests whether subsidiaries that enter the new sector have high human capital compared to the other subsidiaries of the same firm. The dependent variable is a dummy variable that takes the value one if the entry in sector  $n$  at time  $t$  is made through subsidiary  $f$ , zero if the entry is not made through subsidiary  $f$  within firm  $g$ . Entries are identified with sales reported at the 5-digit level of the French SIC. The main independent variable is a subsidiary-level measure of human capital *Internal Human Capital* $_{g,n,t-1}$ . The variable measures the extent to which the workforce of the subsidiary is adapted to the sector of entry (see section 2.3.2). Column (1) and (2) are the baseline specifications. In columns (3) and (4), the measure of Internal Human Capital is calculated at  $t - 2$  and  $t - 3$  respectively. Columns (2) to (4) include the following set of control variables: the number of workers, the subsidiary's cash holdings, amount of tangible assets and value added, the last three variables being scaled by the number of workers in the subsidiary. All models are estimated with firm  $\times$  sector of entry  $\times$  year fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Build})_{f,n,t}$			
	(1)	(2)	(3)	(4)
Internal HC $_{f,n,t-1}$	0.021*** (0.004)	0.024*** (0.004)		
log(Number of employees) $_{f,t-1}$		0.018*** (0.003)	0.017*** (0.003)	0.017*** (0.003)
Cash holdings/#workers $_{f,t-1}$		0.308** (0.153)	0.320** (0.161)	0.308* (0.159)
Fixed assets/#workers $_{f,t-1}$		0.078*** (0.026)	0.079*** (0.027)	0.086*** (0.029)
Value added/#workers $_{f,t-1}$		0.080** (0.032)	0.071** (0.034)	0.062* (0.035)
Internal HC $_{f,n,t-2}$			0.024*** (0.004)	
Internal HC $_{f,n,t-3}$				0.023*** (0.004)
Firm-Entry-Year FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	No
$R^2$	0.229	0.235	0.236	0.234
Observations	362089	362089	316760	267060

**Table 2.10. Human Capital and Workforce Adjustment**

*Source:* matched employer-employee dataset, ownership links dataset, EAE survey. *Sample:* Subsidiaries through which internal entries are realized during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the link between internal human capital and the variation of the number of workers. The main variable in the first three columns is the growth rate of the number of workers (computed as in Davis and Haltiwanger (1992)). The variation is computed within the intervals  $[t-1;t+1]$ ,  $[t-1;t]$ , and  $[t;t+1]$  with  $t$  the year of entry in the new sector. In the last column, the main variable is the simple difference of *Internal Human Capital* $_{g,n,t}$  taken between  $t$  and  $t + 1$ . The main independent variable is a subsidiary-level measure of human capital *Internal Human Capital* $_{g,n,t-1}$ . The variable measures the extent to which the workforce of the subsidiary is adapted to the sector of entry (see section 2.3.2). Control variable include the following set of control variables: the number of workers, the subsidiary's cash holdings, amount of tangible assets and value added, the last three variables being scaled by the number of workers in the subsidiary. All models include sector of origin  $\times$  entry  $\times$  year fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\Delta$ Number of workers $_f$			$\Delta$ Internal HC $_f$
	$[t-1 ; t+1]$ (1)	$[t-1 ; t]$ (2)	$[t ; t+1]$ (3)	$[t ; t+1]$ (4)
Internal HC $_{f,n,t-1}$	-0.013* (0.007)	0.002 (0.001)	-0.016** (0.008)	-0.105*** (0.032)
Origin-Entry-Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
$R^2$	0.639	0.102	0.659	0.422
Observations	19219	26010	19219	16091

**Table 2.11. Human Capital, Diversification, and Local Labor Market Tightness**

*Source:* SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the period 2010-2014 (Pole Emploi started collecting the list of occupations in short supply in 2010). The table reports OLS estimates and tests the effect of tight local labor markets for key occupations on the relationship between human capital and the type of diversification strategy. The dependent variable is a dummy variable that takes the value one if the entry of firm  $g$  in sector  $n$  at time  $t$  is made through an acquisition, zero if the entry is made internally. The main independent variable is a firm-level measure of human capital  $Internal\ Human\ Capital_{g,n,t-1}$ . The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2).  $LLM\ Tightness_{n,z,t}$  is the sum of occupations in short supply in the local labor market  $z$  at time  $t - 1$ , weighted by the occupation  $\times$  sector  $\times$  year fixed effects, scaled by the number of occupations present in sector  $n$  at (see equation (2.4.2)). Terciles of LLM tightness are included in columns (1) and (2). Models in columns (3) to (5) are estimated on subsamples of the dataset split by terciles of LLM tightness. France is divided into 348 local labor markets. Control variables include the number of workers in logarithm as well total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. All models include sector origin  $\times$  sector of entry  $\times$  year fixed effects, as well as labor market fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$				
Tercile of LLM Tightness $_{n,z,t-1}$ :	All	All	$\leq P33$	$\geq P33$ and $\leq P66$	$\geq P66$
	(1)	(2)	(3)	(4)	(5)
Internal HC $_{g,n,t-1}$	-0.016*** (0.005)	-0.008** (0.004)	-0.003 (0.004)	-0.013** (0.006)	-0.016*** (0.004)
2nd tercile of Tightness $_{n,z,t-1}$	0.007** (0.003)	0.006** (0.003)			
3rd tercile of Tightness $_{n,z,t-1}$	0.014*** (0.005)	0.012*** (0.004)			
2nd t. Tightness $_{n,z,t-1} \times$ Int. HC $_{g,n,t-1}$	-0.005** (0.002)	-0.004** (0.002)			
3rd t. Tightness $_{n,z,t-1} \times$ Int. HC $_{g,n,t-1}$	-0.005** (0.002)	-0.004* (0.002)			
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes
$R^2$	0.418	0.424	0.525	0.501	0.377
Observations	28598	28957	7953	7796	8067

**Table 2.12. Local Labor Market Tightness and the Value of Building**

*Source:* SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the period 2010-2014 (Pole Emploi started collecting the list of occupations in short supply in 2010). The table reports OLS estimates and tests the effect of tight local labor markets for key occupations on the relationship between human capital and performance on the sector of entry. The dependent variable  $\log(\text{Sales})_{g,n,z,t}$  is the logarithm of sales realized by firm  $g$  in sector  $n$  the year of entry  $t$ . The variable is defined only for Build entries, i.e., when  $1(\text{Buy})_{g,n,z,t} = 0$ . The main independent variable is a firm-level measure of human capital  $\text{Internal Human Capital}_{g,n,t-1}$ . The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2).  $\text{LLM Tightness}_{n,z,t}$  is the sum of occupations in short supply in the local labor market  $z$  at time  $t - 1$ , weighted by the occupation  $\times$  sector  $\times$  year fixed effects, scaled by the number of occupations present in sector  $n$  at (see equation (2.4.2)). Terciles of LLM tightness are included in columns (1) and (2). Models in columns (3) to (5) are estimated on subsamples of the dataset split by terciles of LLM tightness. France is divided into 348 local labor markets. Control variables include the number of workers in logarithm as well total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. All models include sector origin  $\times$  sector of entry  $\times$  year fixed effects, as well as labor market fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\log(\text{Sales})_{g,n,t}$				
Tercile of LLM Tightness $_{n,z,t-1}$ :	All	$\log(\text{Sales})_{g,n,t}$			
	(1)	(2)	$\leq$ P33 (3)	$\geq$ P33 and $\leq$ P66 (4)	$\geq$ P66 (5)
Internal HC $_{g,n,t-1}$	0.101*** (0.021)	0.094*** (0.028)	0.076 (0.052)	0.048 (0.051)	0.116* (0.060)
2nd tercile of Tightness $_{n,z,t-1}$		0.004 (0.036)			
3rd tercile of Tightness $_{n,z,t-1}$		-0.009 (0.049)			
2nd t. Tightness $_{n,z,t-1} \times$ Int. HC $_{g,n,t-1}$		0.008 (0.029)			
3rd t. Tightness $_{n,z,t-1} \times$ Int. HC $_{g,n,t-1}$		0.012 (0.030)			
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
$R^2$	0.498	0.498	0.569	0.505	0.488
Observations	27760	27760	7635	7512	7786

**Table 2.13. Diversification, Human Capital and Firing Costs**

*Source:* SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the effect of firing costs on the relationship between human capital and the type of diversification strategy. The dependent variable is a dummy variable that takes the value one if the entry of firm  $g$  in sector  $n$  at time  $t$  is made through an acquisition, zero if the entry is made internally. The main independent variable is a firm-level measure of human capital  $Internal\ Human\ Capital_{g,n,t-1}$ . The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2).  $Permanent\ workers\ (\%)_{g,t-1}$  measures the fraction of permanent to temporary workers in firm  $g$ 's workforce at  $t - 1$ . Terciles of  $Permanent\ workers\ (\%)_{g,t-1}$  are included in columns (1) and (2).  $Length\ case\ settlements_{z,t-1}$  measures the average length of local labor case settlements in jurisdiction  $z$  at time  $t - 1$ . There are 140 jurisdictions over the sample period. Terciles of  $Length\ case\ settlements_{z,t-1}$  are included in columns (3) and (4). Control variables include the number of workers in logarithm as well total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. All models include sector origin  $\times$  sector of entry  $\times$  year fixed effects, as well as labor market fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Build})_{f,n,t}$			
	(1)	(2)	(3)	(4)
Internal HC $_{g,n,t-t}$	-0.003** (0.002)	-0.002 (0.001)	-0.010*** (0.003)	-0.010*** (0.003)
2nd tercile of Permanent workers $(\%)_{g,t-1}$	0.001 (0.002)	0.002 (0.002)		
3rd tercile of Permanent workers $(\%)_{g,t-1}$	0.002 (0.002)	0.003 (0.002)		
2nd t. Permanent workers $(\%)_{g,t-1} \times \text{Int. HC}_{g,n,t-1}$		-0.003*** (0.001)		
3rd t. Permanent workers $(\%)_{g,t-1} \times \text{Int. HC}_{g,n,t-1}$		-0.003** (0.001)		
2nd tercile of Length case settlements $_{z,t-1}$			-0.001 (0.002)	-0.000 (0.002)
3rd tercile of Length case settlements $_{z,t-1}$			-0.003 (0.003)	-0.002 (0.003)
2nd t. Length case settlements $_{z,t-1} \times \text{Int. HC}_{g,n,t-1}$				-0.002** (0.001)
3rd t. Length case settlements $_{z,t-1} \times \text{Int. HC}_{g,n,t-1}$				-0.004** (0.002)
Controls	Yes	Yes	Yes	Yes
Origin-Entry-Year FE	Yes	Yes	Yes	Yes
Jurisdiction FE	No	No	Yes	Yes
$R^2$	0.420	0.421	0.515	0.517
Observations	63722	63722	87341	89650

### 2.11.1. Theoretical Appendix

#### Predictions 2.2.2 and 2.2.2

When  $\gamma > (\sigma - 1)/\sigma$ , the difference  $\Delta(c_I, c_E, c_A) = \Pi^{Build}(c_I, c_E) - \Pi^{Buy}(c_I, c_E, c_A)$  is a decreasing function of  $c_I$ . Moreover,  $\lim_{c_I \rightarrow 0} \Delta(\cdot) = F > 0$ , and

$$\lim_{c_I \rightarrow \infty} \Delta(c_I) = F + K \left( c_E^{-\frac{\gamma}{\gamma-1}} \right)^{\frac{(\gamma-1)(\sigma-1)}{\gamma}} - K \left( c_E^{-\frac{\gamma}{\gamma-1}} + c_A^{-\frac{\gamma}{\gamma-1}} \right)^{\frac{(\gamma-1)(\sigma-1)}{\gamma}}, \quad (2.16)$$

which is negative if  $F$  is small enough relative to the marginal labor costs of workers in the target firm. This ensures the unique existence of  $c_I^* > 0$  such that for any  $c_E$  and  $c_A$ ,  $c_I > c_I^*$  implies  $\Delta(\cdot) = 0$ . This proves prediction 2.2.2. Moreover,  $\Delta(\cdot)$  is a decreasing function of  $c_E$ , thus  $\Delta(\cdot) = 0$  for a lower threshold  $c_I^*$  when  $c_E$  is large, implying prediction 2.2.2.

#### Micro-foundation of the fixed cost: Endogenous restructuring costs

We propose a micro-foundation for  $F$ , based on the cost of laying off some workers in the acquired firm. We denote  $\phi$  as the marginal cost of layoff,  $L_A^{before}$  as the labor input of the acquired firm *before* the acquisition, and  $L_A^{after}$  as the labor input *after* the merger has occurred. We assume the layoff cost is proportional to the distance between  $L_A^{before}$  and  $L_A^{after}$ :

$$F(L_A^{before}, L_A^{after}) = \phi(L_A^{before} - L_A^{after}). \quad (2.17)$$

Given the standard monopolistic competition framework, output  $Y$  is equal to  $K(\sigma - 1)c^{-\sigma}$ . Since  $Y = \mathcal{L}$ , we can write the labor input of the acquired firm before the acquisition as

$$L_A^{before} = c_A^{-\sigma} K(\sigma - 1). \quad (2.18)$$

After the target is acquired, the acquiring firm minimizes the total labor cost across the three worker pools  $i \in \{I, E, A\}$ , equal to  $\sum_i c_i L_i^{after}$ , subject to the production function (2.1). The first-order conditions of this problem can be rewritten as

$$L_i^{after} = \frac{c_i^{-\frac{1}{1-\gamma}}}{\sum_i c_i^{-\frac{\gamma}{1-\gamma}}} \cdot \sum_i c_i L_i^{after}, \quad (2.19)$$

so that after the acquisition, the labor input coming from the acquired firm is

$$L_A^{after} = c_A^{-\sigma} K(\sigma - 1) \left( 1 + \left( \frac{c_E}{c_A} \right)^{-\frac{\gamma}{1-\gamma}} + \left( \frac{c_I}{c_A} \right)^{-\frac{\gamma}{1-\gamma}} \right)^{\left( \frac{1-\gamma}{\gamma} \right) (\sigma-1) - 1}. \quad (2.20)$$

Comparing (2.18) and (2.20), we have  $L_A^{after} \leq L_A^{before}$ . It is clear from (2.20) that  $F(L_A^{before}, L_A^{after})$  is decreasing in  $c_I$  and  $c_E$ . Recall that  $\Delta(c_I, c_E, c_A)$  is also a decreasing function of  $c_I$  and  $c_E$ . We now have  $\lim_{c_I \rightarrow 0} \Delta(c_I) = \phi L_A^{before} > 0$ , and

$$\begin{aligned} \lim_{c_I \rightarrow \infty} \Delta(c_I) = & K \left( c_E^{-\frac{\gamma}{\gamma-1}} \right)^{\frac{(\gamma-1)(\sigma-1)}{\gamma}} - K \left( c_E^{-\frac{\gamma}{\gamma-1}} + c_A^{-\frac{\gamma}{\gamma-1}} \right)^{\frac{(\gamma-1)(\sigma-1)}{\gamma}} \\ & + \phi c_A^{-\sigma} K(\sigma-1) \left( 1 - \left( 1 + \left( \frac{c_E}{c_A} \right)^{-\frac{\gamma}{1-\gamma}} \right)^{\left( \frac{1-\gamma}{\gamma} \right)(\sigma-1)-1} \right), \end{aligned} \quad (2.21)$$

$\Delta(\cdot)$  keeps the same properties as before. It is strictly decreasing, positive when  $c_I = 0$ , and negative when  $c_I \rightarrow \infty$  if (2.21) is negative, i.e., if  $\phi$  is not too large. Therefore, the testable predictions stated in the main text remain unchanged.

### Optimal marginal costs (Equation (2.8))

Given a minimum threshold  $\underline{h}_{io}$  and the random match quality  $h \geq 1$  following a Pareto distribution with cf  $\Psi(h)$ , each worker occupation  $o \in \mathcal{O}$  supplies the following amount of human capital:

$$H_{io} \equiv N_i a_{io} m_o \int_{\underline{h}_{io}}^{\infty} h d\Psi(h) = \frac{N_i a_{io} m_o k \underline{h}_{io}^{1-k}}{k-1}, \quad (2.22)$$

with  $i \in \{I, E, A\}$ .

We write the Lagrangian of the minimization of (2.6) subject to (2.7) and (2.22):

$$\begin{aligned} \mathcal{L} = & N_i \left[ \sum_{o \in \mathcal{O}} a_{io} w_o \underline{h}_{io}^{-k} + f c_i \right] \\ & + \sum_{o \in \mathcal{O}} \mu_o \left[ H_{io} - \frac{N_i a_{io} m_o k \underline{h}_{io}^{1-k}}{k-1} \right] \\ & + \lambda_{io} \left[ L_i - \left( \sum_{o \in \mathcal{O}} H_{io}^\theta \right)^{1/\theta} \right]. \end{aligned}$$

We obtain the first-order conditions

$$\mu_o = \frac{w_o}{m_o \underline{h}_{io}} \quad (\text{wrt } \underline{h}_{io}) \quad (2.23)$$

$$c_i = \sum_{o \in \mathcal{O}} \mu_o H_{io} \quad (\text{wrt } N_i) \quad (2.24)$$

$$\frac{w_o}{m_o \underline{h}_{io}} = \lambda H_{io}^{\theta-1} \left( \sum_{o \in \mathcal{O}} H_{io}^\theta \right)^{\frac{1}{\theta}-1} \quad (\text{wrt } H_{io}). \quad (2.25)$$

Plugging (2.23) into (2.25), then rearranging and summing over  $o$ , we obtain

$$C_i = \left[ \sum_{o \in \mathcal{O}} \left( \frac{a_{io} m_o^k w_o^{1-k}}{L_i^k (N_i k)^{-1} (k-1)} \right)^{\frac{\theta}{\beta}} \right]^{\frac{\beta}{\theta(1-k)}}. \quad (2.26)$$

We guess that  $C_i = c_i L_i$ . Using the definition of total costs  $C_i$  given by (2.6) we have that

$$\begin{aligned} c_i L_i &= N_i \left( \sum_{o \in \mathcal{O}} a_{io} w_o h_{io}^{-k} + f c_i \right) \\ &= \sum_{o \in \mathcal{O}} \left( \frac{k-1}{k} \right) \frac{m_o}{w_o h_{io}} H_{io} + N_i f c_i \\ &= \left( \frac{k-1}{k} \right) c_i L_i + N_i f c_i, \end{aligned} \quad (2.27)$$

where we use of Equation (2.22) on the second row and Equation (2.25) on the third row. It follows naturally from (2.27) that  $N_i = \frac{L_i}{f k}$ . Now, plugging  $C_i = c_i L_i$  into (2.26) and using  $N_i = \frac{L_i}{f k}$ , we find that the optimal unit labor cost function for each labor market  $i \in \{I, E, A\}$  is:

$$c_i = \left[ \sum_{o \in \mathcal{O}} \left( \frac{a_{io} m_o^k w_o^{1-k}}{f(k-1)} \right)^{\frac{\theta}{\beta}} \right]^{\frac{\beta}{\theta(1-k)}}. \quad (2.28)$$

### The relative wage share for each worker occupation (Equation (2.9))

Putting together the first-order conditions (2.24) and (2.25) and taking the sum over  $o \in \mathcal{O}$ , we obtain

$$\frac{\frac{w_o}{m_o h_{io}} H_{io}}{C_i} = \frac{H_{io}^\theta}{\sum_{o \in \mathcal{O}} H_{io}^\theta}, \quad (2.29)$$

so that the share of each worker occupation in total costs is equal to the share of human capital supplied by that worker occupation.

We normalize input from each labor market at  $L_i = 1$ , which implies  $\sum_{o \in \mathcal{O}} H_{io}^\theta = 1$ . Plugging (2.22) into (2.29) and using  $N_i = \frac{1}{f k}$ , we obtain

$$h_{io} = \left( \frac{a_{io}}{f(k-1)} \right)^{\frac{1-\theta}{\beta}} w_o^{\frac{1}{\beta}} m_o^{-\frac{\theta}{\beta}} C_i^{-\frac{1}{\beta}}, \quad (2.30)$$

where  $\beta \equiv \theta + \theta(1-k)$ .

From Equation (2.27), we have that the costs of conducting interviews, expressed in labor units, are equal to  $N_i f c_i = \frac{c_i}{k}$ . This implies that for any number of workers  $\tilde{A}_{io}$  to work in an occupation, firms must hire a larger number of workers  $A_{io} =$

$\frac{k}{k-1}\tilde{A}_{io}$  because some workers will be conducting interviews. By assumption, the number of workers in occupation  $o \in \mathcal{O}$  hired writes  $\tilde{A}_{io} = N_i a_{io} (1 - \Psi(h_{io})) = N_i a_{io} h_{io}^{-k}$ . Finally, using (2.30) and  $N_i = \frac{1}{f^k}$ , we find that the total number of workers in occupation  $o$  hired to produce one unit of output is

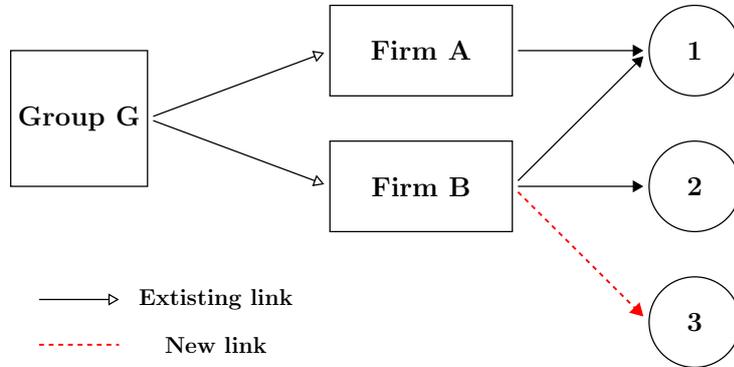
$$A_{io} = \left( \frac{a_{io}}{f(k-1)} \right)^{\frac{\theta}{\beta}} w_o^{-\frac{k}{\beta}} m_0^{\frac{k\theta}{\beta}} C_i^{\frac{k}{\beta}}. \quad (2.31)$$

It follows that relative share of the wage bill that goes to a given occupation  $o$  can be expressed as in Equation (2.9).

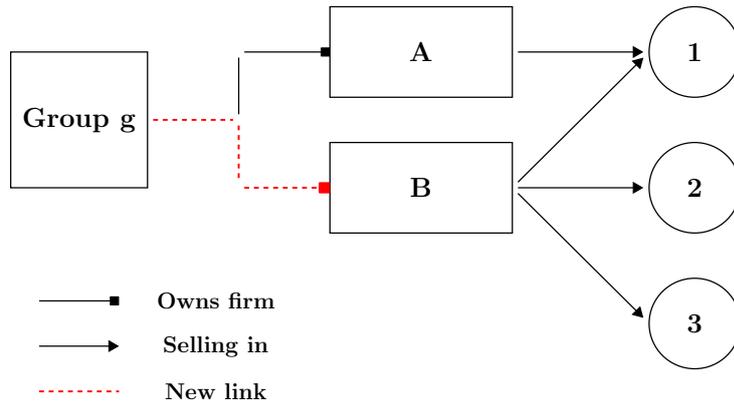
2.11.2. Additional Tables and Figures

Figure 2.A1: Identification of Build and Buy Entries

(a) “Build” entry in sector 3 through entity B



(b) “Buy” entry in sector 2 and 3 through the acquisition of B



**Table 2.A1. Explanatory Power of Equation (2.11)**

*Source:* Matched employer-employee dataset. *Sample:* Firms with more than 20 employees included in the matched employer-employee dataset over the 2003-2007 and 2009-2014 periods. This table gives the adjusted  $R^2$  of the regression of the logarithm of the share of the total wage bill by occupation for a given firm in a given year on various fixed effects. Column 1 corresponds to the estimation that we use to rank occupations by sector and year (ie, equation 2.11).

	(1)	(2)	(3)	(4)	(5)	(6)
Occupation-Sector-Year FE	Yes	Yes	No	No	No	No
Firm-Year FE	Yes	No	Yes	No	No	No
Occupation FE	No	No	No	Yes	No	No
Firm FE	No	No	No	No	Yes	No
Year FE	No	No	No	No	No	Yes
$R^2$	0.661	0.503	0.448	0.178	0.475	0.005
Observations	6262541	6308510	6483176	6526045	6524067	6526045

**Table 2.A2. Top 5 Occupations within Sector**

*Source:* matched employer-employee dataset for the period 2003-2014. *Sample:* Universe of full-time French workers who earn at least 1,000 euros a year. This table reports the top five occupation  $\times$  sector  $\times$  year fixed effects estimated following equation (2.10). Occupation  $\times$  sector  $\times$  year fixed effects capture the average share of the wage bill that goes to an occupation within a sector a given year. The higher the fixed effects the more important the occupation for the production process of a given sector.

## Panel A. Pharmaceutical preparations (2013)

Occupation	Estimated fixed effects
Technicians in production and control quality	0.99
Chemists, operators and skilled workers	0.96
R&D engineers and executives	0.61
Sales managers in SMEs	0.31
Sales representatives and technicians	0.29

## Panel B. IT consultancy activities (2013)

Occupation	Estimated fixed effects
Computer science R&D engineers and executives	2.34
IT project manager	1.63
CEOs of service companies (1-49 workers)	0.99
CEOs of commercial companies (1-49 workers)	0.92
IT support engineers and executives	0.82

## Panel C. Manufacture of motor vehicles (2013)

Occupation	Estimated fixed effects
Mechanical qualified assemblers in series	0.99
Technicians specialized in mechanics and metal work manufacturing and quality control	0.85
CEOs of companies (50-499 workers)	0.40
R&D engineers specialized in mechanics and metal work	0.40
Unskilled workers in assembly lines of metal work	0.14

**Table 2.A3. Top 10 sectors of entry in 2013**

The table reports the number of entries for sectors with the largest number of entries, by entry type in 2013. Build entries (Panel A) are identified as sales reported into a new sector (source: ESA survey). Buy entries (Panel B) are identified when firms realize diversifying acquisitions (source: SDC Platinum and Bureau van Dijk Zephyr). Sectors refer to an industry at the 4-digit level of the French SIC.

Panel A. Build entries		
Sector of origin	Sector of entry	N. Pairs origin × entry
Retail sale in stores with food or tobacco	Manufacture of bread, pastry and cakes	880
Retail sale in stores with food or tobacco	Renting and leasing of cars and light motor vehicles	565
Retail sale in stores with food or tobacco	Agents involved in the sale of food	488
Retail sale in stores with food or tobacco	Retail sale of automotive fuel in specialized stores	395
Retail sale in stores with food or tobacco	Retail sale via mail order or Internet	338
Sale of cars and light motor vehicles	Maintenance and repair of motor vehicles	332
Retail sale in stores with food or tobacco	Renting or leased real estate	296
Freight transport by road	Other transportation support activities	295
Freight transport by road	Warehousing and storage	278
Freight transport by road	Maintenance and repair of motor vehicles	272
Panel B. Buy entries		
Sector of origin	Sector of entry	N. Pairs origin × entry
Computer consultancy activities	Computer programming activities	14
Activities of head offices	Wholesale of machinery and equipment	8
Software publishing	Computer programming activities	8
Computer consultancy activities	Software publishing	7
Computer consultancy activities	Business and other management consultancy activities	6
Activities of head offices	Wholesale of other household goods	6
Engineering activities	Programming activities	6
Wholesale of machinery and equipment	Manufacture of lifting and handling equipment	5
Computer consultancy activities	Wholesale of computers, peripheral equipment and software	5
Computer consultancy activities	Other business support service activities	5

**Table 2.A4. Top 10 Occupations in Short Supply**

*Source:* French unemployment agency. *Sample:* Occupations' supply and demand of 348 local labor markets in 2013. This table reports the top 10 occupations that are reported as being in short supply the most often among local labor markets. The last column gives the percentage of local labor markets in which the occupation is reported as being in short supply. Occupations in short supply are identified by the French national employment agency as occupations for which (i) the ratio of job offers over job applications is high (ii) surveyed employers forecast that it will be difficult to fill posted offers.

Rank	Occupation	% of local labor markets
1	Kitchen staff	77.0
2	Machining equipment operators	58.3
3	Butchers	51.7
4	Metal workers	50.0
5	Nurses	48.9
6	Technical and commercial relation managers	48.3
7	Bakers	48.0
8	Car mechanics	48.0
9	Catering staff	48.0
10	Machining equipment maintenance workers	46.2

**Table 2.A5. Robustness Check: Main Results with 3-Digit Sector Level**

*Source:* SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and analyses the role of human capital on the type of diversification strategy. The dependent variable  $1(\text{Buy})_{g,n,t}$  is a dummy variable that takes the value one if the entry of firm  $g$  in sector  $n$  at time  $t$  is made through an acquisition, zero if the entry is made internally. Entries are identified with sales reported at the 3-digit level of the French SIC. The main independent variable is a firm-level measure of human capital *Internal Human Capital* $_{g,n,t-1}$ . The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2). Control variables include the number of workers in logarithm as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include Sector Origin  $\times$  Entry  $\times$  Year fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$1(\text{Buy})_{g,n,t}$				
	(1)	(2)	(3)	(4)	(5)
Internal HC $_{g,n,t-1}$	-0.018*** (0.004)	-0.013*** (0.004)			-0.003* (0.002)
$\log(\#\text{workers})_{g,t-1}$		0.010*** (0.002)	0.010*** (0.002)	0.011*** (0.002)	-0.052*** (0.012)
Cash holdings/ $\#\text{workers}_{g,t-1}$		0.091* (0.053)	0.090 (0.062)	0.093 (0.066)	0.162 (0.137)
Fixed assets/ $\#\text{workers}_{g,t-1}$		-0.048 (0.034)	-0.056 (0.036)	-0.049 (0.036)	-0.195** (0.089)
Value added/ $\#\text{workers}_{g,t-1}$		0.052 (0.053)	0.040 (0.053)	0.028 (0.060)	-0.421*** (0.116)
Internal HC $_{g,n,t-2}$			-0.015*** (0.004)		
Internal HC $_{g,n,t-3}$				-0.017*** (0.005)	
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No	Yes
$R^2$	0.175	0.182	0.184	0.184	0.545
Observations	81180	81117	70964	59206	63750

**Table 2.A6. Robustness Check: Alternative Measure of Local Labor Market Tightness**

*Source:* SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the period 2010-2014 (Pole Emploi started collecting the list of occupations in short supply in 2010). The table reports OLS estimates and tests the effect of tight local labor markets on the relationship between human capital and the type of diversification strategy. The dependent variable is a dummy variable that takes the value one if the entry of firm  $g$  in sector  $n$  at time  $t$  is made through an acquisition, zero if the entry is made internally. The main independent variable is a firm-level measure of human capital  $Internal\ Human\ Capital_{g,n,t-1}$ . The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2).  $LLM\ Tightness_{n,z,t}$  is the sum of occupations in short supply in the local labor market  $z$  at time  $t - 1$ , scaled by the number of occupations present in sector  $n$  (see equation (33)). Terciles of LLM tightness are included in columns (1) and (2). Models in columns (3) to (5) are estimated on subsamples of the dataset split by terciles of LLM tightness. France is divided into 348 local labor markets. The set of control variables include the number of workers, total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. All models include sector origin  $\times$  sector of entry  $\times$  year fixed effects, as well as labor market fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable: Level of LLM Tightness $_{n,z,t-1}$ :	All		$1(Buy)_{g,n,t}$		
	(1)	(2)	$\leq P33$ (3)	$\geq P33$ and $\leq P66$ (4)	$\geq P66$ (5)
Internal HC $_{g,n,t-1}$	-0.015*** (0.004)	-0.009** (0.004)	-0.000 (0.002)	-0.011** (0.005)	-0.019*** (0.007)
2nd tercile of Tightness $_{n,z,t-1}$	0.007* (0.004)	0.006* (0.004)			
3rd tercile of Tightness $_{n,z,t-1}$	0.012** (0.005)	0.012** (0.005)			
2nd t. Tightness $_{n,z,t-1} \times$ Int. HC $_{g,n,t-1}$	-0.003** (0.002)	-0.003 (0.002)			
3rd t. Tightness $_{n,z,t-1} \times$ Int. HC $_{g,n,t-1}$	-0.007** (0.003)	-0.005* (0.003)			
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	Yes
LLM FE	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes
$R^2$	0.418	0.424	0.498	0.543	0.400
Observations	28957	28957	7897	7778	7882

**Table 2.A7. Robustness Check: Are Public Acquirers Special?**

*Source:* French unemployment agency. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests whether the coefficients of the different control variables depend on whether the firm is public or not. Public firms are those that include at least one publicly listed subsidiary within the firm. The dependent variable  $\mathbb{1}(\text{Buy})_{g,n,t}$  is a dummy variable that takes the value one if the entry of firm  $g$  in sector  $n$  at time  $t$  is made through an acquisition, zero if the entry is made internally. Entries are identified with sales reported at the 5-digit level of the French SIC. The main independent variable is a firm-level measure of human capital  $\text{Internal Human Capital}_{g,n,t-1}$ . The variable measures the extent to which the workforce of the firm is adapted to the sector of entry (see section 2.3.2). Control variables include the number of workers in logarithm as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include Sector Origin  $\times$  Entry  $\times$  Year fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. \*, \*\*, and \*\*\* mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Internal HC <sub><math>g,n,t-1</math></sub>	-0.010*** (0.003)	-0.010*** (0.003)	-0.011*** (0.003)	-0.010*** (0.003)	-0.011*** (0.003)	-0.010*** (0.003)
Cash holdings/#workers <sub><math>g,t-1</math></sub>	0.044 (0.045)	-0.008 (0.046)	0.037 (0.045)	0.022 (0.045)	0.019 (0.045)	0.041 (0.046)
Fixed assets/#workers <sub><math>g,t-1</math></sub>	-0.022 (0.028)	-0.022 (0.028)	-0.029 (0.029)	-0.019 (0.028)	-0.021 (0.028)	0.002 (0.028)
Value added/#workers <sub><math>g,t-1</math></sub>	0.049 (0.035)	0.050 (0.035)	0.047 (0.035)	0.012 (0.034)	0.044 (0.035)	0.006 (0.035)
log(#workers) <sub><math>g,t-1</math></sub>	0.010*** (0.002)	0.010*** (0.002)	0.010*** (0.002)	0.009*** (0.002)	0.008*** (0.002)	0.008*** (0.002)
Public $\times$ Int. HC <sub><math>g,n,t-1</math></sub>	-0.009 (0.017)					-0.004 (0.018)
Public $\times$ Cash/#workers <sub><math>g,t-1</math></sub>		0.945*** (0.242)				-0.454 (0.359)
Public $\times$ Fixed A./#workers <sub><math>g,t-1</math></sub>			0.250* (0.129)			-0.655*** (0.214)
Public $\times$ VA/#workers <sub><math>g,t-1</math></sub>				0.809*** (0.129)		0.957*** (0.301)
Public $\times$ log(#workers) <sub><math>g,t-1</math></sub>					0.009*** (0.002)	0.006* (0.004)
Origin-Entry-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
$R^2$	0.206	0.208	0.207	0.211	0.210	0.212
Observations	76296	76296	76296	76296	76296	76296

## 2.12. Description of Variables

Variables	Description
<u>Dependent variables</u>	
$\mathbb{1}(\text{Build})_{f,n,t}$	Dummy variable that takes the entry into sector $n$ is made through a firm $f$ within the firm $g$ , and zero if the entry is <i>not</i> made through the firm $f$ . This variable is constructed only for the subsample of Build entries for which we can identify precisely the entity responsible for the entry. <i>Source</i> : ownership links dataset, ESA survey.
$\mathbb{1}(\text{Buy})_{g,n,t}$	Dummy variable that takes the value one if a firm $g$ enters a sector $n$ at year $t$ through an acquisition, and zero if the entry is made internally by building on preexisting resources. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Build entries are identified using reported sales from the ESA survey at the 4-digit of the French SIC. <i>Source</i> : SDC Platinum, BvD Zephyr, Tax files, Ownership links dataset, ESA survey.
Build entries	Number of occurrences in which new sales are reported in a new sector. <i>Source</i> : ESA survey.
Buy entries	Number of occurrences in which firms realize diversifying acquisitions. <i>Source</i> : SDC Platinum and Bureau van Dijk Zephyr.
$\Delta \text{workers}_f [t-i;t+j]$	Growth rate of the number of workers computed following Davis and Haltiwanger (1992): $\Delta \text{workers}_{t-i,t+j} = \frac{\# \text{Workers}_{t-i,t+j} - \# \text{Workers}_{t-i,t+j-1}}{0.5 * (\# \text{Workers}_{t-i,t+j} + \# \text{Workers}_{t-i,t+j-1})}$ <p>. The growth rate ranges between -2 and 2. The variation is computed between <math>[t-1;t+1]</math>, <math>[t-1;t]</math> and <math>[t;t+1]</math> with <math>t</math> the year of entry in the new sector. <i>Source</i>: Matched employer-employee dataset.</p>
$\Delta \text{Internal Human Capital}_{g,n} [t;t+1]$	Simple difference of Internal Human Capital $_{g,n,t}$ between $t$ and $t + 1$ with $t$ the year of entry in the new sector. <i>Source</i> : Matched employer-employee dataset.
$\log(\text{Sales})_{g,n,t}$	logarithm of sales realized by firm $g$ in sector $n$ the year of entry $t$ . <i>Source</i> : Tax files.
<u>Independent variables</u>	
$\mathbb{1}(\text{No Top 10})_{g,n,t-1}$	Dummy variable that takes the value 1 if none of the 10 most important occupation for the sector $n$ of entry are present in the workforce of firm $g$ at time $t - 1$ , and zero if there are. <i>Source</i> : Matched employer-employee dataset.
Internal Human Capital $_{g,n,t-1}$	Sum of (exponentiated) occupation $\times$ sector $\times$ year fixed effects present both in the workforce of firm $g$ at time $t - 1$ and in the sector of entry $n$ . The measure is scaled by the number of occupations in firm $g$ . Fixed effects are estimated on the full French matched employer-employee dataset and are retrieved from a regression that takes the share of the wage bill that goes to a given occupation within a given sector as dependent variable. It can be interpreted as the fit quality of a firm workforce to a sector prior to diversification. <i>Source</i> : Matched employer-employee dataset.
Internal Human Capital $_{g,n,t-i}$	Lagged measure of <i>Internal Human Capital</i> at time $t - 2$ or $t - 3$ . The variable is constructed using the same method as Internal Human Capital $_{g,n,t-1}$ <i>Source</i> : Matched employer-employee dataset.
Internal Human Capital (weighted) $_{g,n,t-1}$	Variable constructed using the same method as Internal Human Capital $_{g,n,t-1}$ , except the sum of occupation $\times$ sector $\times$ year fixed effects is weighted by the number of employees by occupation in the firm. <i>Source</i> : Matched employer-employee dataset.

*Continued next page*

### Description of Variables (continued)

Variables	Description
Internal Human Capital (without CEO) $_{g,n,t-1}$	Variable constructed using the same method as Internal Human Capital $_{g,n,t-1}$ , except the CEO occupations are excluded from the sum of occupations. <i>Source:</i> Matched employer-employee dataset.
Internal Human Capital (plant-level) $_{g,n,t-1}$	Variable constructed using the same method as Internal Human Capital $_{g,n,t-1}$ , except the fixed effects are estimated at the plant-level instead of the firm level. <i>Source:</i> Matched employer-employee dataset.
<u>Control variables</u>	
Cash holdings/N.workers $_{g,t-1}$	Total cash holdings of firm $g$ at time $t - 1$ , scaled by the number of workers in the firm. <i>Source:</i> Tax files, Ownership links dataset, Matched employer-employee dataset.
Fixed assets/N.workers $_{g,t-1}$	Total value of fixed assets held by firm $g$ at time $t - 1$ , scaled by the number of workers in the firm. <i>Source:</i> Tax files, Ownership links dataset, Matched employer-employee dataset.
Length case settlements $_{z,t-1}$	Average length of local labor case settlement in jurisdiction $z$ at time $t - 1$ . There are 140 jurisdictions in France over the sample period. <i>Source:</i> Matched employer-employee dataset, French Ministry of Justice.
LLM tightness $_{n,z,t-1}$	Sum over all occupations in short supply in the local labor market $z$ at time $t$ of the exponentiated occupation $\times$ sector $\times$ year fixed effects for sector $n$ , scaled by the number of occupations observed in sector $n$ at time $t$ (see equation 2.4.2). Occupations in short supply in a LLM are identified by the French national employment agency Pole Emploi as occupations for which (i) the ratio of job offers over job applications is high (ii) surveyed employers forecast that it will be difficult to fill posted offers. France is divided into 348 different local labor markets. <i>Source:</i> Matched employer-employee dataset, Pole Emploi.
log(N.workers) $_{g,t-1}$	Logarithm of the number of workers in firm $g$ at time $t - 1$ . <i>Source:</i> Matched employer-employee dataset, Ownership links dataset.
New geographical zone $_{g,t-1}$	A dummy that indicates whether the firm enters a new geographical zone at time $t$ . We use the "region" as definition of the geographical zone (France is divided in 25 regions over the sample period). A firm enters a new geographical zone if either (i) the entering subsidiary is created at time $t$ and located in a new region, (ii) the entering subsidiary opens a plant in a new region at time $t$ or (iii) the entering subsidiary is acquired at time $t$ and operates in a region in which the firm was not present at $t - 1$ . <i>Source:</i> Matched employer-employee dataset, Ownership links dataset, SIRENE.
Value added/N.workers $_{g,t-1}$	Total value added generated by firm $g$ at time $t - 1$ , scaled by the number of workers in the firm. <i>Source:</i> Tax files, Ownership links dataset, Matched employer-employee dataset.
Permanent workers (%) $_{g,t-1}$	Fraction of workers employed in permanent contracts in firm $g$ at time $t - 1$ ; <i>Source:</i> Matched employer-employee dataset.
Product market distance $_{g,n,t-1}$	Distance between the sectors in which firm $g$ operates at $t - 1$ and the sector of entry $n$ . The distance ranges from 0 to 1. A value of 0 (resp. 1) indicates that the sales of firm $g$ are perfectly correlated (resp not correlated) at time $t - 1$ to the sales of the firms operating in the sector of entry $n$ . The product market distance is adapted from Bloom, Schankerman and Van Reenen (2013). See section 2.4.2 for more details. <i>Source:</i> Ownership links dataset, Tax files, ESA survey.

*Continued next page*

### Description of Variables (continued)

Variables	Description
Vertical integration $_{g,n,t-1}$	Dummy variable that is equal to one if the sector of origin of the firm (as measured in the I-O OECD classification) sources more than 5% of its inputs from the sector of entry (as measured in the I-O OECD classification). The variable is not defined when the sector of origin and the sector of entry belongs to the same item of the I-O OECD classification. The method follows Fan and Goyal (2006)'s measure of vertical relatedness. <i>Source:</i> Ownership links dataset, Tax files, OECD 1999 French Input-Output table.
<u>Other variables</u>	
Entry sales $_{g,n,t}$	Total sales reported by firm $g$ in sector $n$ at time $t$ . <i>Source:</i> Tax files
$\mathbb{1}$ (1-year survival) $_{g,n,t+1}$	Dummy equal to one if firm $g$ reports sales in sector of entry $n$ one year after the entry $n$ at time $t$ . <i>Source:</i> Tax files
Investment bucket $_{g,n,t}$	Decile of total capital expenditures realized by firm $g$ to enter sector $n$ . Investment is measured in the case of build entry as the total amount of investment realized by the firm(s) of the firm that entered in sector $n$ at time $t$ . In the case of acquisitions, investment is set equal to the amount of fixed assets of the firm that has been acquired. The investment buckets are used as fixed effects to compare firms within the same investment bucket. <i>Source:</i> Tax files
Sales bucket $_{g,n,t-1}$	Deciles of realized sales by the firm $g$ in sector $n$ at time $t$ . The sales buckets are used as fixed effects to compare firms within the same entry sale bucket. <i>Source:</i> Tax files
Serial acquirers $_g$	Serial acquirers are firms that enter more than one sector by acquisition during the time period. <i>Source:</i> SDC Platinum, BvD Zephyr, Ownership links dataset, ESA survey
Shifting firms $_{g,t}$	Shifting firms are firms that decrease their activity in a preexisting sector while entering sector $n$ . They are identified with the growth rate of sales between $t - 1$ and $t$ in each sector in which firms were operating at $t - 1$ that is negative and greater than 100%, 50% and 25% in absolute value. We take the firm-level minimum of sectoral growth rates. <i>Source:</i> ESA survey, Ownership links dataset.
Public firm $_g$	Dummy variable that is equal to one if at least one entity within the firm is a publicly listed company, and zero otherwise. <i>Source:</i> Bureau van Dijk Amadeus

## Chapter 3

# Are Investors Aware of Ownership Connections?

\*

with Edith Ginglinger (Université Paris Dauphine) and Luc Renneboog (Tilburg University)

---

\*This chapter benefited from comments and discussions with Paul Beaumont, Frans Belot, Fabio Braggion, Riccardo Calcagno, Fabio Castiglionesi, Gilles Chemla, Peter Cziraki, Eric de Bodt, Mark Deloof, Jocelyn, Evans, Frank de Jong, Olivier Dessaint, Joost Driessen, Bart Dierynck, Joost Driessen, Pascal Dumontier, Zsuzsanna Fluck, Ivalina Kalcheva, Yuexin Li, Alberto Mancini, Sophie Manigart, Roni Michaely, Mario Milone, Klaas Mullier, Manju Puri, Emanuele Rizzo, David Robinson, Oliver Spalt, Christophe Spaenjers, Cara Vansteenkiste, Michael Weisbach, Marius Zoican as well as the seminar participants at Tilburg University, Université Paris-Dauphine, VU Amsterdam Tinbergen Institute, Ghent University, University of Mannheim, 3rd HEC PhD workshop, AFFI 2018, FMA European conference 2018, EFMA 2018, the Corporate Finance Day at the University of Antwerp and the Paris December Finance conference 2018. This paper received the Larry Lang Best Corporate Finance Paper at the EMFA 2018.

## Abstract

We examine market reactions to earnings announcements within a parent-subsi-dary ownership structure. Parents' investors react to all announcements within the group either immediately or with delay, whereas subsidiaries' investors only react to their own firm's announcements, ignoring predictive information released by the parent. Multiple announcements within a group lead to *enhanced transparency* for parents' investors. In contrast, subsidiaries' investors appear unaware of ownership links, and behave as *inattentive* investors. Inattention is worsened by geographical diversifica-tion and indirect ownership, but is explained by strategic timing of the disclosure of earnings surprises, day-of-the-week effect, internal capital markets, or synergy-related explanations across industries.

### 3.1. Introduction

Recent evidence in the literature suggests that public information is not automatically impounded into stock prices. Investors' limited attention may be inattentive to earnings news released by the firm itself or by companies that are connected to it through economically significant contractual links induces delayed stock price response.<sup>2</sup> When the firm is organized as a network of a parent and several subsidiaries<sup>3</sup>, relevant inside information about the group has multiple issuers. Disclosure by one company within the network can be directly relevant for affiliated entities as equity stakes represent channels through which earnings float. In this paper, we examine how shareholders respond to earnings announcements by the various entities of business groups. We investigate whether these structures lead to enhanced transparency as investors receive more detailed information coming from different entities or to more opacity when investors are unable to comprehend the connections between the announcing entities.

To examine information flows within groups, we identify parent-subsidiary ownership structures where both parent and subsidiary are publicly listed and their sets of shareholders only partially overlap. Throughout the paper, we will use the term subsidiaries for firms in which a parent owns at least a 20% stake and are following IFRS standards on consolidation. Thus, a parent company's stock can be regarded as a weighted portfolio of listed and privately-owned subsidiaries, where the weights are determined by ownership stakes in subsidiaries and the subsidiaries' relative sizes. At least once a year, the listed parent and subsidiary companies are required to make public and separate announcements of their earnings. By studying the market reactions to the release of unanticipated information, we can identify whether investors are able to see through the complexity of the group structure, or whether corporate connections induce investor unawareness or inattention. Indeed, the fact that a related listed parent and subsidiary have to a large extent a distinct set of investors, and that both entities release separate earnings information enables us to examine how investors react to information disclosure of the affiliated entity.

We first document that there is relevant information for the subsidiary's (parent's) shareholders in the earnings announcement of the parent (subsidiary) by highlighting a positive relationship between announcements' informational contents from

---

<sup>2</sup> On investor inattention to earnings news disclosed by their own firm, see for example Bernard and Thomas (1989, 1990); Engelberg and Gao (2011); DellaVigna and Pollet (2009); Hirshleifer, Lim and Teoh (2009). On investors' inattention to news from connected firms, see Cohen and Frazzini (2008) for customer-supplier relationships, Cao, Chordia and Lin (2016) for strategic alliances and Massa and Žaldokas (2017) for co-ownership.

<sup>3</sup> We define corporate networks as a group of legally independent firms connected by ownership links. These networks are largely prevalent in Asia (Claessens, Djankov and Lang, 2000), continental Europe (Faccio and Lang, 2002), but also exist in the U.S. (see for example Holderness (2007), who discredits the myth of diffuse ownership in the U.S.).

both firms. We then distinguish among three cases based on the timing of earnings information release: (i) the parent and subsidiary announce on the same day, (ii) the parent announces first, and (iii) the subsidiary announces first. Market efficiency predicts that investors of both parents and subsidiaries fully and immediately incorporate earnings information unanticipated by the market, the anticipated part of the information being already priced. When the earnings announcements do not coincide, the first company to announce is expected to also convey predictive information about the affiliated entity that announces second. Consequently, unanticipated information released by the first announcer should not only trigger an immediate share price reaction for the first announcer but also for the second one. Hence, investors are expected to perceive the ownership links, entailing that the surprise earnings of all announcing companies belonging to the group increase the amount of information available to investors (*enhanced information hypothesis*).

The alternative hypothesis is that investors are unaware of the ownership links and do not react, neither immediately nor with delay to the earnings announcements of the affiliated entity (*inattention hypothesis*). A variation of this hypothesis is that investors are *heterogeneously* inattentive: some are aware of the ownership links, others are not, inducing delays in processing information released by affiliated entities. We consider the delayed stock price reaction to earnings surprises, measured by the post-earnings announcement drift (Already in the first footnote) as an indication of investor inattention (Bernard and Thomas, 1989, 1990; DellaVigna and Pollet, 2009; Hirshleifer, Lim and Teoh, 2009).

Our analysis is based on a sample of 15,117 ownership link-year observations, corresponding to 2,181 unique (direct or indirect) parent-subsidiary links in 75 countries over the period 2000-2015 and yields the following results. First, we find that when the parent and the subsidiary release earnings surprises on the same day, investors of both firms strongly and immediately react to the announcements. However, these pooled announcements do not enable us to identify the information source to which each set of investors reacts. Second, when the parent company releases its earnings information prior to subsidiary, we find that parent's investors react both to their own surprise earnings announcement and to the subsidiary's announcement that takes place at a subsequent point in time. This implies that the parent's investors infer that the subsidiary's announcement contains additional information that was not priced yet at the parent's initial announcement of the aggregated information of the whole group. In contrast to a parent's investors, the subsidiary's investors only react to the subsidiary's announcement, ignoring the ex-ante and hence predictive information released at the parent level. This suggests that the subsidiary's investors are generally unaware of the ownership relation between subsidiary and the parent firm, and fail to observe that the entity is part of a group. Third, when the sub-

subsidiary announces its earnings surprise first, we observe that both the subsidiary's and parent's investors immediately incorporate this information in the share prices. However, they only do so partially as the share prices keep adjusting in the period after the subsidiary's announcement to fully incorporate the news (reflected in the PEAD). The immediate reaction to a firm's own announcements accounts for 50-66% of the total reaction, whereas the immediate reaction to the other firm's announcements merely represents 0 to 15% of the total reaction. When the subsidiary announces first, not all parent's investors seem to be aware of the ownership connections.

We explore three potential explanations for our findings. First, investors may have a blind spot in the sense that they do not perceive that companies are part of a group. Hence, investors do not react to the announcements of an affiliated company as they fail to observe the internal structure of the corporate group. When the group consist of entities located in different countries, operating in different industries, or whose corporate names do not reveal any connection, it may be harder for investors to comprehend that these entities are all part of the same group and to process information released by its various members. This would be especially the case when information is disclosed by group members in which the investors do not hold a direct ownership stake. The theoretical argument is based on Merton (1987)'s model in which the investors are not aware of the entire universe of securities and obtain information on a small number of stocks, leading to neglected stocks. In our setting, this would mean that investors collectively do not perceive the group as a whole and only consider the entity in which they have directly invested. We find that investors are less inattentive when a subsidiary is located in the same country, is directly owned, and when the parent and subsidiary share part of a corporate name, all characteristics that may increase investor awareness of the group structure.

Second, there are several reasons why subsidiary's investors, as opposed to parent's investors, may be more subject to inattention. Investors are more likely to obtain a broad picture of a complex group when they invest in the apex company of the group rather than in an entity within the corporate network. Furthermore, from the point of view of a listed subsidiary's outside investors, the parent company's news, even if it is disclosed first, may be less informative because other subsidiaries' performance may blur that of the listed subsidiary. Hence, understanding how the consolidated news relates to the individual entities of the group may require more sophisticated analysis. Finally, absence of or delayed investor reaction, which we interpret as investor unawareness, may be driven by limits of arbitrage due to illiquid stocks and transaction costs (see for example Bhushan (1994); Ng, Rusticus and Verdi (2008)). (i.e. many subsidiaries may have a lower free float given the presence of a major shareholder, the parent, (see for example Bolton and Von Thadden (1998);

Maug (1998)). In all our models, we control for the Amihud (2002) stock illiquidity measure, which makes it unlikely that illiquidity effects are the main drivers of our findings.

Third, we investigate whether shareholder characteristics may explain our results. Smart investors may be more aware of ownership links and better able to use information released by different entities of the group. We focus our analysis on institutional investors, who are more likely to be sophisticated and to initiate large trades.<sup>4</sup> To capture institutional investor heterogeneity, we include in our models the percentage of capital owned by mutual funds and active investors (private equity and hedge funds). Following the literature on common ownership to identify common investors (Azar, Schmalz and Tecu, 2018; Gilje, Gormley and Levit, 2018), we find that having common active investors who hold equity in both parent and subsidiary leads to a stronger initial reaction to the subsidiary's earnings announcements but does not change the post-earnings announcement drift. As Ke and Ramalingegowda (2005) emphasize, exploiting PEAD requires institutional investors to actively trade stocks. They find no evidence that dedicated and quasi-indexing institutions exploit PEAD. In contrast, transient institutions, characterized by high portfolio turnover, trade to exploit PEAD especially in firms with low transaction costs. One potential explanation for our results of persistent PEAD could be that business groups, due to their complexity and relative opacity, do not attract enough transient institutional investors to trade away PEAD. To sum up our results, institutional investors do not seem to be smarter at understanding complex firm structures, with the exception of active investors owning shares in both the parent and subsidiary companies.

Fourth, we explore whether financial analysts, who are commonly viewed as sophisticated investors, are able to help investors understand the links between the different entities in the group, by updating their forecast after the first announcement to include related firm released information. We find that analysts who update their forecast in the 7 days (30 days) following the first announcement (and before the second announcement, made by the affiliated firm) represent only 2.4% (5.5%) of all analysts making predictions for the second announcer. Our results suggest that financial analysts provide limited assistance to investors to better understand multiple announcements in corporate groups, previous literature on financial analysts' abilities (Abarbanell and Bernard, 1992; Bradshaw, Richardson and Sloan, 2001).

---

<sup>4</sup> A large body of research suggests that institutional investors are better informed or have an advantage in processing publicly available information around earnings announcements, see for example Walther (1997); Bartov, Radhakrishnan and Krinsky (2000); Bhattacharya (2001). Among institutional investors, some may be more sophisticated than others. Examining the net buying activity of investors in response to earnings announcement, Battalio and Mendenhall (2005) find that large traders use a complete information set that incorporates analysts' forecasts, whereas small traders ignore earnings signals based on analysts' forecasts and respond to signals of a less accurate time-series model.

We conduct several tests to address endogeneity issues and alternative explanations. First, it could be that the decision to announce first or jointly may be endogenous to the quality of the news. Managers of the parent and/or subsidiary may try to steer different share price reactions by taking advantage of the announcement timing (Stein, 1989). In the specific case of listed parent-subsidiary structures, financial communication calendars can be (de)synchronized depending on the quality of the news announcement (e.g. Begley and Fischer, 1998; Bagnoli, Kross and Watts, 2002; Doyle and Magilke, 2009; Boulland and Dessaint, 2017; Johnson and So, 2018). We do find that the announcement timing is related to the quality of the earning surprise and to characteristics of the subsidiary and the parent, but we do not find that investors' reactions are driven by the timing of who releases earnings surprise information first and on whether the earnings surprise is positive or negative. Therefore, we confirm that moral hazard is not the main friction at play to explain investor behavior.

Second, when the subsidiary announces first, we interpret the delayed market reaction as resulting from investor unawareness. This finding may not just result from complexity induced by ownership links but from releasing information on days when inattention is usually high. The traditional inattention literature examines investor distraction with regard to information timing of stand-alone firms, such as the effect of information announcements on Friday (??), busy days (Hirshleifer, Lim and Teoh, 2009; Frederickson and Zolotoy, 2015), and busy hours (Michaely, Rubin and Vedralshko, 2013). In all our models, we control for year, month, and day of the week fixed-effects.

Third, while we interpret the absence of a subsidiary's investor reaction to the surprise earnings announcement by the parent who announces first, as the result of unawareness of the ownership link, an alternative explanation could be that the subsidiary's investors fear tunneling by the parent entity (Johnson et al., 2000; Dyck and Zingales, 2004) and perceive the reported accounting information as uninformative (Fan and Wong, 2002). Additional tests do not support that our findings could be affected by expropriation of subsidiary earnings by the parent.

Fourth, a related idea is that an internal capital market within the group may exist and enables capital transfers to financially constrained firms (Stein, 1997; Rajan, Servaes and Zingales, 2000). We find that, when a parent announcing first has negative earnings, investors react more favorably to the subsequently announced positive earnings surprise from the subsidiary, and this effect is amplified when the parent's growth opportunities are larger than the subsidiary's. This test suggests that investors may value the existence of internal capital markets.

Fifth, we test the robustness of all our findings to any consideration related to the endogenous formation of the groups (e.g. vertical integration, diversification). We

include pair (parent-subsidiary) industry (country) fixed effects to account for unobserved complementarities and synergies between parent companies and subsidiaries operating in different industries (countries), and we confirm that these effects do not affect the way investors react to information released by their company or by the affiliated entity.

Our paper is related to several strands of literature. First, we add to the literature examining the importance of ownership structures to explain the magnitude of the post-earnings announcement drift. PEAD is one of the most persistent documented anomalies. There are plenty of results on the impact of shareholder types on the PEAD. For instance, Kaniel et al. (2012) find that informed trading by individuals is responsible, at least in part, for PEAD, especially for smaller firms. Ke and Ramalingegowda (2005) report that only institutions with high portfolio turnover rates exploit PEAD. Prado, Saffi and Sturgess (2016) find that higher ownership concentration stocks tend to have a lower speed of adjustment to earnings announcements and a bigger PEAD. Our findings suggest that the presence of institutional investors (categorized by type, nationality, or size) does not reduce the PEAD. In contrast, we find that common active investors (hedge funds and private equity) owning shares in both parent and subsidiary companies contribute to speed up the adjustment of prices to earnings announcements.

Second, we contribute to the literature on inattention limited attention to information within complicated firms (the *dark side* of such firms).<sup>5</sup> Cohen and Lou (2012) compare standalone and conglomerate firms subject to the same information shock. They argue that investors' limited processing capacity leads to a significant delay in impounding information into share prices of conglomerate firms, generating return predictability. Barinov, Park and Yildizhan (2018) find that an increase in firm complexity leads to larger post-earnings announcement drifts. Huang (2015) reaches the same conclusion looking at multinational corporations relatively to focused US firms. As conglomerates and multinational corporations are non-exclusive forms of complex firms, we consider in this paper both dimensions of complexity: sectoral and geographic diversification. In addition, by examining internal ownership connections within firms, we open the black box of complex firms, and we show that investors' processing capacity depends on the characteristics of the links between the entities of the group.

Third, we add new evidence that such complex firms' structure can also turn out to be beneficial for investors (*bright side* of complex firms). The literature has identified settings where within-conglomerate information sharing can generate value: for instance, Massa and Rehman (2008) find that mutual funds operated by

---

<sup>5</sup> In addition, several papers examine valuation issues of complex firms, see for example Slovin and Sushka (1997); Lamont and Thaler (2003); Laeven and Levine (2007).

financial conglomerates have superior performance, arguably because information is shared by their banking division. Anjos and Fracassi (2015) argue that conglomerate firms have an informational advantage relative to focused firms because they have better access to business-relevant information, especially if they operate in more “central” industries relative to the global industry network. Our analysis suggests that connected firms yield higher transparency that could be beneficial for all the investors in the various entities of the group, but is actually only picked up by some types of investors.

Fourth, we contribute to the literature on inattention to information from connected firms. Ramnath (2002) finds that investors underreact to the earnings reports of rivals within an industry. Cohen and Frazzini (2008) and Madsen (2017) show that stock prices do not fully incorporate news related to a firm’s principal customers. Cao, Chordia and Lin (2016) examine the impact of information released by one partner in a strategic alliance on the share price of the other partner. They document a share price underreaction to information release by the other partners regardless of whether the information is positive or negative. As these papers examine firms related through different types of external links but do not analyze corporate relations within ownership networks, our paper tries to fill this gap by providing evidence of inattention to news released inside a business group.

Fifth, we add to the literature examining the importance of ownership structures to explain the magnitude of the post-earnings announcement drift. PEAD is one of the most persistent documented anomalies. There are plenty of results on the impact of shareholder types on the PEAD. For instance, Kaniel et al. (2012) find that informed trading by individuals is responsible, at least in part, for PEAD, especially for smaller firms. Ke and Ramalingegowda (2005) report that only institutions with high portfolio turnover rates exploit PEAD. Prado, Saffi and Sturgess (2016) find that higher ownership concentration stocks tend to have a lower speed of adjustment to earnings announcements and a bigger PEAD. Our findings suggest that the presence of institutional investors (categorized by type, nationality, or size) does not reduce the PEAD. In contrast, we find that common active investors (hedge funds and private equity) owning shares in both parent and subsidiary companies contribute to speed up the adjustment of prices to earnings announcements.

The remainder of the paper is organized as follows. In Section 2, we describe the sample selection and give descriptive statistics of the main variables. We report our results in Section 3 and the results from a set of robustness tests in Section 4. Section 5 concludes.

## 3.2. Sample selection

### 3.2.1. Ownership links

We start our data collection by retrieving shareholder information for all (currently and formerly) listed companies around the world from Bureau van Dijk's Orbis database. We find 360,000 ownership links between a public company and a public (downstream) company.<sup>6</sup> Still, most of these ownership links are participation stakes held by financial institutions including insurance companies (45% of all of the above links) or mutual funds (25%) and these equity stakes are small with an average (median) of 4.52% (0.56%) of the equity. Some of the equity stakes in Orbis are not given in a numerical format, which is why we decode them: we replace a percentage with a leading "<", or ">" by the percentage after the symbol plus or minus 0.1%; we eliminate possible signs that precede percentages: "\_", "?", or "Â"; the "WO" codes (wholly owned) are replaced by 98.01%; "MO" (majority owned) by 50.01% (because according to the international accounting standards practice, majority ownership is at least 50% plus one share and the smallest stake reported by BvD is at two decimals, 0.01%); "CQP1" (50% plus 1 share) by 50.01%; "BR" (branch) by 50.01%; "JO" (jointly owned) by 50%; "NG" (negligible) by 0.01%; and "n.a" (not available) and "-" (not significant) by zero.

Our aim is to identify investor reactions to a credible signal emitted by a related company, *i.e.* a subsidiary that directly or indirectly significantly contributes to the parent's earnings. *laporta2000* define a large owner as a legal entity that directly or indirectly controls at least 10% of the voting rights. *claessens2000* use a 20% cutoff to study concentrated ownership structures. We follow the literature and retain the ownership links with a percentage equal to or above 20%. Since 2005, there has been a strong push for harmonization of accounting standards and principles with the mandatory adoption of International Financial Reporting Standards (IFRS) for publicly traded firms, which largely coincides with U.S. GAAP. Both U.S. GAAP and IFRS require parent companies to consolidate controlled subsidiaries. IFRS standards require the parent to consolidate the entity if there is *de facto* control, which is interpreted as the parent owning a stake of 20% or more (see Appendix B).

Our 2015 cross-sectional sample of ownership links comprises 14,353 subsidiary-parent relations involving 20,616 listed companies. We drop 4,537 links where ISIN codes are missing. We expand the sample going back 16 years (2000-2015) and obtain a panel of 54,917 link-year observations based on ownership links of publicly

---

<sup>6</sup> Bureau van Dijk's Orbis database provides owner and subsidiary links for more than 40 million public and private companies. The data are collected from different data sources including the SEC Edgar files for US listed companies, firms' annual reports, firms' websites, and direct solicitations. Orbis relies on a network of 77 local data providers to collect international ownership data.

listed parent companies that effectively directly or indirectly control at least 20% percentage of the equity of publicly listed subsidiaries.<sup>7</sup>

### 3.2.2. Earnings surprises

We collect earnings announcement dates, realized earnings per share, and analysts' earnings forecasts, as provided by the I/B/E/S U.S. and International files.<sup>8</sup> We follow the accounting and finance literature by defining earnings surprises as the difference between the announced earnings and the analysts' forecasts from the period prior to the announcement. Following *della Vigna 2009*, we take each analyst's most recent forecast prior to the announcement provided that the forecast is between 180 and 3 days before the announcement (to avoid recent forecasts being affected by leakage of information on realized earnings). Our earnings forecast is the median of all analysts' forecasts.<sup>9</sup> We scale the difference between the realized earnings and the median analyst forecast by the share price taken five trading days prior to the announcement.<sup>10</sup> Thus, our estimate of the earnings surprise for firm  $i$  on day can be written as:

$$Surprise_{i,t} = \frac{(actual\ earnings_{i,t} - median\ forecast_{i,t,[-180+;-3]})}{price_{i,t-5}}$$

The variable *Top Two Quantiles* <sub>$i$</sub> , which is the independent variable of interest, is defined following *della Vigna 2009*:

$$Top\ Two\ Quantiles_{i,t} = \begin{cases} 1, & Surprise_{i,t} \in \{Q10; Q11\} \\ 0, & Surprise_{i,t} \in \{Q1; Q2\} \end{cases}$$

The variable *Surprise* <sub>$i$</sub>  distribution is split into 11 quantiles  $Q$ , with  $Q6$  being the quantile with a *Surprise* <sub>$i$</sub>  closed to zero,  $[Q7; Q11]$  the quantiles with positive *Surprise* <sub>$i$</sub> , and  $[Q1; Q5]$  the quantiles with negative *Surprise* <sub>$i$</sub> .

As we work with a global sample, we convert all quantities to USD by means of daily exchange rates from Datastream. We delete the observations with extreme earnings surprises (absolute value greater than one). We focus on the annual earn-

<sup>7</sup> We correct the data for potential mistakes; e.g. we delete the link-year observations prior to the IPO year, and after the delisting and full takeover year. IPO and delisting dates are collected from BvD Amadeus and Datastream. Takeovers dates are collected from BvD Zephyr as it has common identifiers with the Orbis database. We also retrieve all historical ownership links available in the Orbis Historical files related to companies involved in the sample of cross sectional links.

<sup>8</sup> We link the Bureau van Dijk Orbis information to Datastream and I/B/E/S databases using the ISIN identifier.

<sup>9</sup> Considering the median analyst forecast gives no weight to analysts that perform poorly at issuing earnings forecasts.

<sup>10</sup> An alternative methodology scales forecast error by the standard deviation of earnings forecasts, but this necessitates at least two analysts following a company, which is not always the case, especially for subsidiaries.

ings announcement because the practice of quarterly earnings announcements is not universally mandatory, and companies subject to IFRS around the world are required to announce their earnings on an annual basis Hung, Li and Wang (2014).<sup>11</sup> In an international context, most studies find that annual earnings announcements are informative, especially for firms in countries with higher quality earnings and with stronger investor protection institutions (DeFond, Hung and Trezevant, 2007), and after firms cross-listed in the U.S. (Bailey, Karolyi and Salva, 2006).

### 3.2.3. Investors' reactions

The stock return at an earnings announcement represents the change in a firm's valuation induced by investors' buying and selling transactions triggered by the difference in earnings relative to expectations. We compute cumulative abnormal returns for different windows at the date  $j$  of parent's and subsidiary's earnings announcement for each set of investors - where  $i = \{p, s\}$ ,  $p$  stands for the parent and  $s$  for the subsidiary. We download daily returns from Datastream and denote  $r_{i,t}$  as the returns of the share of a company  $i$  on day  $t$ . We calculate cumulative abnormal returns,  $BHAR_{i,j} [j; j + T]$ , over a  $[i; i + T]$  window as buy-and-hold returns:

$$BHAR_{i,j} [i; i + T] = \prod_{t=i}^{i+T} (1 + r_{i,t}) - 1 - \widehat{\beta}_{i,t} \left[ \prod_{t=i}^{i+T} (1 + r_{m,t}) - 1 \right],$$

where  $r_{m,t}$  is the daily market portfolio return.  $\widehat{\beta}_{i,t}$  are obtained by regressing individual returns on the MSCI World 600 index returns for an estimation window  $[-300; -46]$ . We drop the announcements for which we have less than 40 days of stock price data for the estimation period.

For each pair of parent  $p$  and subsidiary  $s$  in each year  $t$  of the sample period, we study two sets of investor reactions at two earnings announcement dates, ' $p$ ' and ' $s$ ', yielding a total of four reaction-announcement observations in each year.

Our main test is captured by the following equation:

$$BHAR_{i,j} [j; j + T] = \delta Surprise\ Top\ Two\ Quantiles_{i,j} + ' Firm\ Controls_{i,j} + \Theta Link\ Controls_{p,s,j} + \alpha_{p,s} + b_j + \varepsilon_{i,j}$$

<sup>11</sup> As of today, 114 countries have converged to IFRS (see Appendix B for more details). In many countries, the usefulness of mandatory reporting of quarterly earnings has been questioned, as they are believed to strengthen a short-term focus at the expense of the long run. E.g. the Interim Management Statements, introduced in 2007 in the UK, were abandoned in 2014. In 2013, the European Commission amended its Transparency Directive stating that quarterly financial information is not necessary for investor protection.

where the vector  $Firm\ Controls_{i,j}$  comprises the firm characteristics including the log of market capitalization, the log of analyst coverage, the market-to-book ratio, and the Amihud (2002) illiquidity measure.  $Link\ Controls_{p,s,i}$  is a vector of a pair  $(p, s)$  characteristics including the companies' relative market size, percentage of common analysts, percentage of control held by the parent, a dummy variable indicating a direct ownership relation, and dummy variables equaling one if the parent and subsidiary are located in the same country, and have part of their name in common, respectively. We also include industry and time (year, month, day-of-the-week) fixed effects ( $b_i$ ), and in some specifications pair (parent-subsidiary) industry fixed effects<sup>12</sup> or link fixed effects ( $a_{p,s}$ ). The dependent variable  $BHAR_{i,i}$  is calculated by type of investor (*i.e.*  $p$  or  $s$ ), each of which is expected to respond to the surprise earnings announcements of  $p$  or  $s$  (at ' $p$ ' or ' $s$ '). We therefore examine four cases: parents' investor reactions to the parent companies' announcements ( $BHAR_{p,p}$ ), subsidiaries' investor reactions to subsidiaries' announcements ( $BHAR_{s,s}$ ), parents' investor reactions to subsidiaries' announcements ( $BHAR_{p,s}$ ), and subsidiaries' investor reactions to parents' announcements ( $BHAR_{s,p}$ ).

### 3.3. Description of the sample

#### 3.3.1. Geographic breakdown

The geographical distribution of parents and subsidiaries spans 75 countries. We partition these countries into six categories, in addition to the U.S. and Great Britain. The category 'Asia' includes China, Hong Kong, Korea, Indonesia, Japan, Myanmar, Singapore, Philippine, India, Singapore, and Thailand. The category 'Europe' includes Albania, Austria, Belgium, Croatia, Cyprus, Czech Republic, Germany, Denmark, Spain, Finland, France, Ireland, Greece, Hungary, Italy, Lithuania, Luxembourg, The Netherlands, Norway, Poland, Portugal, Romania, Russia, Slovakia, Sweden, Switzerland, and Ukraine. The 'Americas' comprises Canada, Latin America, and Caribbean islands. The group Middle East includes Israel, Turkey, Lebanon, Oman, and Saudi Arabia. AU/NZ stands for Australia and New Zealand.

55% of the business groups around the world are from Asia. Within the Asian groups, 58% are Japanese, about 10% are Chinese (including Hong Kong), 13% are Korean, and the remainder is from India (6%), Singapore (4%), Thailand (4%), Myanmar (4%) and Indonesia (2%). Western continental Europe stands for 16% of all the corporate groups (globally), the U.S. for 11%, the Great Britain for 4% and the rest of the world for 14%.

---

<sup>12</sup> This way, we control for the fact that the different industries can be more strongly or weakly connected through customer or supplier trade flows, as argued by Ahern and Harford (2014).

[Insert table 3.1 here]

### 3.3.2. Descriptive statistics

Table 3.2 reports descriptive statistics of the main explanatory and control variables. Means (medians) of various firm characteristics are reported for the subsamples of parents, subsidiaries, and parent-subsidiary pairs. We provide a complete list of variable definitions in Appendix A. Parents (subsidiaries) are on average followed by 14 (8) analysts. The average (median) size of the subsidiary represents 41% (12%) of the average (median) parent.

Parent company's stocks are more liquid than the subsidiaries', which have on average a lower free float. The average parent (subsidiary) has 25 (13) institutional investors, which control 16% (11%) of the average parent (subsidiary). Institutional investors are identified using Bureau van Dijk Orbis and include mutual funds, pension funds, hedge funds, venture capital and private equity funds, banks and insurance companies that own between 0.1% and 20% of the equity. 38% of the parents and subsidiaries have at least one common institutional investor. On average, they have four common owners.

The average parent-subsidiary structures are geographically focused (72% of the links are in the same country), diversified (58% operate in different industries), and 27% share part of the corporate name. The average ownership stake by the parent in the subsidiary amounts to 49% and the relative market value of the subsidiary is 41%, such as the average subsidiary represents 20% ( $0.49 \times 0.41$ ) of the parent's value, which is economically important enough to expect investors to react to the earnings release of the affiliated entity.

[Insert table 3.2 here]

## 3.4. Empirical results

In this section, we focus first on parents' and subsidiaries' investor reactions to the earnings announcements of their own firms as well as to those of the affiliated companies. We then examine the effect of relative announcement timing on share prices in order to understand when (if at all) information is incorporated into stock prices.

### 3.4.1. Investor Reactions to Earnings Announcements

A parent's earnings announcement contains information on its various operations, including those of its subsidiaries, such that the parent's announcement conveys

information on the business group, relevant to both the parent's and subsidiaries' investors. Similarly, when a subsidiary announces its earnings first, the market receives predictive information about part of a parent's consolidated future earnings.

In table 3.3, we regress investor reactions on announcements of earnings surprises belonging to the top and bottom quantiles of their distribution. The parameter estimate of *Surprise Top Two Quantiles* measures the returns to good news (top two quantiles) relative to bad news (bottom two quantiles) and we expect the coefficient to be positively correlated with the investor reaction. Our empirical setting with two sets of investors and two earnings events per year enables us to study the investor reaction to the earnings announcement for four combinations of investor reaction-firm announcement. We examine the cases when firms announce their earnings on the same day and when either the parent or the subsidiary announces first. For each combination, we examine, controlling for company's characteristics and industry (SIC-2) fixed effects, at the immediate and the delayed response to the event day which we label day 0: windows [0;1] and [2;60], respectively. For models that test investor reactions to the earnings surprise of the affiliated company, we include (i) characteristics of that affiliated company (firm size, analyst coverage, market-to-book (Q), and stock illiquidity), (ii) link (parent-subsidiary) control variables (including the parent's ownership percentage in the subsidiary, the percentage of common analysts, the presence of common institutional investors, and the size of subsidiary relative to the parent's), (iii) pair (parent-subsidiary) industry fixed effects that account for unobserved heterogeneity not only related to these firms' own industries but also to combination of industries wherein the parent and its subsidiary operate and (iv) time-independent characteristics of the links (dummies for a parent and subsidiary being located in the same country, operating in the same industry, or bearing a common name). This specification accounts for the endogenous formation of business groups, driven by unobserved complementarities, synergies, or inadequacies between industries. We include year, month, and day-of-the-week fixed effects, in order to rule out the effects of business cycles, within-year seasonality, day-of-the-week inattention, and report robust standard errors.

Table 3.3 presents the results of the share price reactions: Buy-and-hold abnormal returns at the announcement period [0;1] and the subsequent period [2;60] measuring the post-earnings announcement drift (PEAD). We examine whether parent's and subsidiaries' investors react differently to the announcement of their own firm and the affiliated firm. Columns (1) and (2) report the parent's investor reaction to the parent's own announcement. We find a strong, immediate positive reaction of 1.51%, and a PEAD not significantly different from zero. The parent's investor reaction to its subsidiary's surprise earnings announcements are shown in columns (3) and (4): These investors incorporate the information released by the

subsidiary into the parent company's share price, although some of the information is incorporated with delay (at the 10% level, we find a PEAD of 0.8%). Columns (5) and (6) reveal that the subsidiary's investors react to the subsidiary's earnings announcement. Half of the price reaction takes place immediately and the other half up to 60 trading days after the announcement. The PEAD is economically and statistically significant and robust to the subsidiary's controls.

In sum, our results support the fact that parent's and subsidiary's investors do not react in the same way to information released by the affiliated entity. In contrast to the parent's investors, subsidiary's investors do not incorporate information on the affiliated firm into their share prices. This violation of the efficient market hypothesis may follow from a lack of investor sophistication (inattention) or alternative explanations that we explore further in the next section.

[Insert table 3.3 here]

### **3.4.2. Relative announcement timing and information incorporation**

We dissect our sample into three subsamples based on different announcement timing situations, respectively the case where the parent and its subsidiary release their earnings on the same day, where the parent announces before the subsidiary, and where the subsidiary announces first. For most parents and subsidiaries (about 84%) within a pair, the financial year ends coincide (table 3.4). Of these pairs, only 16% announce their annual earnings on the same day. For 47%, the subsidiary's earnings announcement is scheduled before that of the parent; and for 37%, the subsidiary releases its earnings subsequent to the parent. When the subsidiary is the first announcer, the parent announces on average (median) 13 (7) calendar days afterward. When the parent is the first announcer, the subsidiary's announcement is scheduled 23 days later (median 14 days).

[Insert table 3.4 here]

For the cases where parent and subsidiary do not announce on the same day, we perform four different tests: we measure the parents' and subsidiaries' investor reactions to the announcements of their own firms, and their reactions to the announcements of the affiliated firms. In each setting, we study investor reaction or inattention, and estimate immediate and delayed reactions to the earnings surprises.

#### **Are the earnings reported by affiliated entities correlated?**

Before embarking on how the earnings announcements are received by the various types of investors, we verify whether the size of the earnings reported by parent

and bidder are correlated. We find strong correlations in regressions of the parents' (subsidiaries') earnings surprises (belonging to the top two quantiles of their distribution) on the subsidiaries' (parents') earnings surprises (also in the top two quantiles) when the subsidiary (parent) is announcing first: 0.62 (0.58). Repeating these regressions with the actual earnings of the parents and subsidiaries confirms these strong relations.<sup>13</sup> We then turn to table 3.5 where we study whether these relations are upheld when controlling for link (parent-subsidiary) controls, pair SIC-2 industry fixed effects, and year, month, day-of-the-week fixed effects. In models (1) and (2), we examine the case of the subsidiary releasing its earnings first and we regress the earnings reported by the parent on those released by the subsidiary. We confirm that parent's actual earnings are positively and significantly related to the subsidiary's actual earnings. The parent's earnings surprises (belonging to the Top Two Quantiles) increase by about 20% when the subsidiary earnings surprise also belong to the Top Two Quantiles. In models (3) and (4), we perform the same tests for the parent announcing prior to the subsidiary and find a similar and strong relation.

These results suggest that actual earnings released by the two companies matter for the affiliate company and that the magnitude of the two earnings surprises are positively related. Hence, good news released by the first announcer predicts good news for the second announcing company, suggesting that investors in the second announcing company should infer information about news to be released by the second announcer.

[Insert table 3.5 here]

### **Parent and subsidiary announce on the same day**

table 3.6 reports the average immediate and delayed reactions of parent's and subsidiary's investors to same-day earnings announcements. Both the parent's and subsidiary's investors instantaneously react to the announcement of their respective companies in models (1) and (5). Consistent with market efficiency, the reaction is immediate and there is no post-earnings announcement drift (Models (3) and (7)). The results remain unchanged when we control for firm characteristics such as stock illiquidity, firm size, analyst coverage, and industry and time (year, month, day-of-the-week) fixed effects (see Models (2), (4), (6) and (8)). As the stock price reactions may also reflect the incorporation of information about the affiliated company given that the announcements coincide, we cannot identify whether the stock price reactions are driven by information released by the parent, the subsidiary, or

---

<sup>13</sup> Firms affiliated through ownership connections may also be linked through an exchange of labor, called internal labor markets (Faccio and O'Brien, 2017).

both. Therefore, we turn to the cases where the parent and the subsidiary disclose their earnings at distinct moments in time.

[Insert table 3.6 here]

### **The parent announces first**

**Parent's investor reactions.** When the parent releases its earnings before the subsidiary, the parent's investors could react twice in case the second announcement also contains new or previously unpriced information. We find that parent's investors strongly react to their own earnings surprises, with a statistically significant BHAR of almost 1% on the announcement day (table 3.7, Panel A, Model 1) and that there is no delayed stock price reaction (Model 3). These results are robust to including variables capturing illiquidity, firm size, Tobin's Q, and the number of analysts, link controls (subsidiary-parent relative market value, percentage of control held by the parent, percentage of common analysts, a dummy capturing the type of ownership link (direct or indirect), and dummy variables equal to one if parent and subsidiary are located in the same country, if they share part of a corporate name, and if common institutional investors own a stake in both entities) and fixed effects (firm-industry, pair industry, and year, month, and day-of-the week).

As a parent's earnings reflect the consolidated earnings from its listed and non-listed subsidiaries, the stock price reaction at the announcement is expected to fully incorporate all relevant information. Still, we show that parent's investors do also react to the subsequent release of surprise information by the subsidiary (Model (5)). This result implies that the uncertainty about the drivers of the earnings surprise at the parent level is partly resolved when the subsidiary discloses earnings surprises and that this additional information about parent's earnings disaggregation is still valuable to the parent's investors. Extending the controls in Model (6) with parent firm characteristics, and pair industry fixed effects yields similar results. The parent's investors hence react instantaneously to the disclosure of earnings surprises of both the parent and the subsidiary and there is no evidence of any delayed reaction when the parent is the first announcer (Models (7) and (8)).

News related to complex ownership links may require more effort to process as opposed to news from stand-alone firms. Similarly, news from geographically and operationally diversified firms may also be more difficult to collect and analyze. A subsidiary's earnings surprise announcement sheds light on the breakdown of the parent's performance. The total parent's investor reaction is about 1.51% (=1.17%+0.34%; models (1) and (5), and similar numbers are obtained in models (2) and (6) which combine to 1.55%). Hence, 77% (=1.17%/1.51%) of the total information is processed at the parent's announcement and 23% at the subsidiary's.

These combined returns are close to those triggered by the parent's investor reactions when the earnings surprise announcements are made by the parent and subsidiary on the same day (1.63%, Model (1) in Table 6). The results remain qualitatively the same when we control for the parent's and link characteristics, and pair industry fixed effects (columns (2) and (6)).

To sum up, when the parent releases its earnings first, its investors react both to their own firm's earnings announcements and to those of the subsidiary, which implies that the latter announcement still contains some additional information not yet incorporated in the parent's share price. This finding supports the *enhanced transparency* hypothesis for the parent's investors.

[Insert table 3.7 here]

**Subsidiary's investor reactions.** In panel B of table 3.7, we test whether the subsidiary's investors react to the information released by the parent when it discloses prior to its subsidiary. As subsidiary's earnings are consolidated in the parent's earnings, it would be rational for subsidiary's investors to immediately incorporate earnings surprise information released by the parent into the subsidiary's stock price. The reason why a subsidiary's stock price is to react is that (i) the parent's and subsidiary's earnings are strongly correlated (see above), (ii) the subsidiary's market capitalization is large relative to the parent's (on average 41%), and (iii) the parent's ownership stake in the subsidiary is on average 49%. Therefore, the effect of the announcement on a subsidiary's stock price ought to take the above effects into account multiplied by the probability that the surprise earnings at the consolidated level are partially induced by this specific subsidiary's earnings. However, we do not find any economically or statistically significant subsidiary's price reaction to the parent's announcement – neither immediately (Models (1) and (2)) nor with a delay (Models (3) and (4)). The subsidiary's investors only react to their own firm's earnings announcements.<sup>14</sup> Then, the response is both immediate (2.5% in Models (5) and (6)), and with delay (Models (7) and (8)). In contrast to parent's investors, the subsidiary's investors only react to the subsidiary's announcement, ignoring the predictive information released at the parent level. This lack of stock price reaction implies that the subsidiary's investors fail to see how their entity is embedded in the business group, suggesting that subsidiary's investors are mostly unaware of ownership links.

---

<sup>14</sup> It could be rational for the subsidiary's investors not to react to the prior parent's announcement if the subsidiary's contribution to the parent (relative size x ownership stake) is small. As a robustness test, we therefore exclude the subsidiaries with small contributions and only retain the case with a contribution of at least 20%, 30% and 40%. We still do not find any subsidiary stock price reaction to the parent's announcement. We then also include interaction variables of the above measurements of contribution and the Top Two Quantiles dummy (for the parent), and also do not find statistical significance.

To sum up, in the case where the parent announces first, the parent's investors incorporate information beyond that of the parent into the share prices, which implies that the subsidiary's earnings announcement adds value and enhances the transparency about the parent company. The subsidiary's investors only react to the information of the subsidiary itself and seem to be inattentive towards predictive surprise news from the parent (which supports the inattention hypothesis).

### **The subsidiary announces first**

We now turn to the case where a subsidiary's earnings announcement is scheduled prior to that of the parent. We expect that the subsidiary also conveys predictive information for the parent's investors. The signal would be particularly informative when the subsidiary's earnings constitute a large part of the parent's consolidated earnings. Thus, we expect the parent's investors to react twice: first, at the subsidiary's earnings surprise announcement, and then at the parent's disclosure. We expect the subsidiary's investors to respond to the disclosure of their own firm but not to that of the parent because, in principle, all relevant information for the subsidiary's investors is already released at this first announcement date.

**Parent's investor reactions.** Panel A of table 3.8 reports the parent's investor reaction to the subsidiary's and parent's earnings surprises when the subsidiary announces first. At the subsidiary's announcement, the parent's stock prices immediately react by on average 0.4% (Model (1)), but most of the information is processed with delay over the period [2;60] days as the stock prices then still increase by 2.8% (Models (3) and (4)). At its own (subsequent) announcement, the parent's stock prices immediately react (by 1.2% in Models (5) and (6)), but again keep adjusting over the subsequent period (Models (7) and (8) exhibit a significantly positive PEAD of almost 2%).

Our findings highlight that when the subsidiary announces first, the parent's investors exhibit heterogeneous behaviors: they react with delay to both their own earnings announcements and those of the affiliated company. This suggests that at least some investors are not aware of the ownership connection with the subsidiary, or are unable to swiftly interpret information related to corporate complexity. We investigate these two potential mechanisms below, in table 3.9 and 3.10.

[Insert table 3.8 here]

**Subsidiary's investor reactions.** Panel B of table 3.8 reveals that at a subsidiary's announcement, its stock price on average immediately reacts (by 2.5%) to the unexpected positive news (Models (1) and (2)), and keeps adjusting upwards

over the subsequent period (Models (3) and (4)). Most of the information is hence seeping in the stock prices with a delay. It should be noted that the PEAD is not induced by the second announcement (see robustness tests in Section 4.4). Models (5) to (6) show that the subsidiary's stock price does not immediately respond to the parent's surprise earnings announcement (which occurs at a stage subsequent to those of the subsidiaries) but only do so with a delay.

There are several reasons why unawareness of ownership links may be more severe for a subsidiary's investor. Whereas a parent investor has indirectly invested in the subsidiary whose earnings contribute to the parent's results, a subsidiary's investor has usually not invested in the parent. Moreover, it is probably easier to have a broad picture of a complex group when investing in the head rather than in a part of the group. Parents release earnings that consolidate the earnings of the publicly listed subsidiary and of the privately-owned subsidiaries and divisions for which no public separate earnings announcement is required. Hence, while it may be relatively straightforward to incorporate a subsidiary's information into the share price of the parent, it generally requires more sophisticated analysis to do the inverse and interpret the impact of earnings information of the parent (which comprises information of the network connections) on a subsidiary's share price. From the point of view of a subsidiary's outside investors, although the parent's news is disclosed first and contains predictive information about the subsidiary's earnings, this information may not be easy to disentangle from other entities' performance. Another reason for the absence of a subsidiary's share price response to the parent's earnings announcement may be related to lower liquidity of the subsidiary's stock because of more concentrated ownership and a smaller free float, which may coincide with fewer institutional investors. If a majority of the shareholders in the subsidiary are non-sophisticated investors, the reaction to announcements, especially for complicated firms, may not happen or may be understandably delayed – a point we will revisit in the next section.

Overall, our results highlight that when a subsidiary announces first, the information value seems blurred and more difficult to understand, and hence triggers share price reactions with a delay.

### 3.4.3. Channels of investor unawareness

In this subsection, we focus on parent's investor reaction to the subsidiary's announcement when the subsidiary releases its earnings first, and we investigate whether firm complexity or heterogeneity in investors' sophistication may explain why information is incorporated with delay.<sup>15</sup>

---

<sup>15</sup> We also test the subsidiary's investor reaction to the parent's announcement when the parent releases its earnings first. We find that the absence of subsidiary's investor reaction is not influenced

### 3.4.4. Corporate complexity

Panel A of table 3.9 reports a delay in the parent's stock price reaction to the subsidiary's announcement when the parent and the subsidiary are not located in the same country. Similarly, in Panel B, we find that parent's investors immediately and significantly react to directly owned subsidiaries (while part of the information is also priced later), but when the subsidiary is controlled through several layers of intermediate firms, the parent's investors only react to subsidiary's earnings surprise with delay. The results in Panel A and B suggest that the complexity induced by geographical diversification and by indirect ownership makes part of investors more inattentive to information within the network.

When the parent and subsidiary share part of their corporate names, the link between these firms should be easier to identify. Still, Panel C shows that the reactions to the earnings announcements are similar whatever subsidiaries' name. In Panel D, we distinguish between the cases where the parent controls more or less than 50% of the subsidiary' equity and the results are also similar.

[Insert table 3.9 here]

### 3.4.5. Investor sophistication

Abnormal returns around earnings announcements result from investors modifying their holding positions in reaction to firms' prospects. Investors are more likely to incorporate information when they are sophisticated and professional investors closely following the company. However, even institutions can be passive investors who do not closely manage their portfolio (Appel, Gormley and Keim, 2016). We test this hypothesis by relating the parent's delayed reaction to the subsidiary's release of predictive information in the context of the presence of institutional and common owners. 57% of parents and 50% of subsidiaries have at least one institutional investor. Collectively, institutional investors have a 16% stake in the parent, and own 11% of the subsidiary. Parent's investors who also invest in the subsidiary hold on average 4.5% in the subsidiary.

In table 3.10, panel A, besides controlling for strategic ownership (held by families and governments), we also control for institutional investors which we partition into three categories: mutual funds, active investors (private equity funds, venture capital funds, and hedge funds), and banks and insurance companies (including other financial firms). We find that our previous results remain qualitatively unchanged. In table 3.10, panel B, we examine the effect of common institutional owners (who own shares in both parent and subsidiary) on the immediate and delayed reactions

---

by the characteristics of the corporate network and the level of subsidiary's investor sophistication. This is why we do not show the tables equivalent to tables 9-13 for this specific case.

to subsidiaries' announcement. Common institutional owners are likely to be more aware of ownership ties. As in Panel A, we divide common institutional owners into three categories, and find that common institutional owners do not affect stock price reactions. Still, active investors' common ownership (private equity funds, venture capital funds, and hedge funds) is positively related to abnormal returns at the announcement, but does not reduce the post earnings announcement drift. Distinguishing between local institutional investors and US institutions does not affect the results. One potential explanation for our results of persistent PEAD could be that business groups, due to their complexity and relative opacity, do not attract enough transient institutional investors to trade away PEAD.

Overall, our analysis suggests that investors have heterogeneous abilities to detect ownership connections. Institutional investors do not seem to be smarter investors, the exception being common active owners, who contribute to accelerate incorporation of information at the subsidiary's earnings announcement.<sup>16</sup>

[Insert table 3.10 here]

## 3.5. Robustness Checks

### 3.5.1. Endogenous strategic announcement timing

The strategic timing literature posits that managers can exploit investor inattention by scheduling their earnings announcements. If the managers of parents and subsidiaries know that the immediate and delayed stock price responses will differ based on which of the affiliated companies first announces positive or negative surprises, they may set up relative announcement timing strategies and coordinate their announcements. Managers could schedule the announcement of good-news-subsidiaries first, and bad-news-subsidiaries after the parent's announcement. By means of Heckman sample selection models, we test possible strategic timing by examining whether stock price reactions to earnings surprises (stage 2 in the models of table 3.11) are affected by announcement timing (stage 1). The results shown in table 3.11 are robustness tests on the parent's (columns 1-2) and subsidiary's (columns 3-4) investor reactions conditional on the relative announcement timing (scheduled on the same versus on different days). Columns (5)-(7) show the parent's investor reactions to the announcement of the subsidiary conditional on the subsidiary announcing prior to the parent (relative to the inverse case).

---

<sup>16</sup> In table 3.8, we reported that there was no subsidiary's reaction to the parent's announcement when this occurred first. When we added to that analysis the institutional ownership of the subsidiary, the findings do not change. In the average subsidiary, mutual funds, activists, and banks/insurers own respectively, 4.4%, 1.1%, and 6.0%; for the subsidiaries with a large contribution to the parent (20% or more), the numbers augment to 7.2%, 1.6% and 9.1%.

Our findings reveal that the choice to schedule a parent's and subsidiary's earnings surprise announcements on the same day does not affect the way the parent's and subsidiary's investors react to the announcements (in columns (2) and (4), respectively): both instantaneous reactions are significantly positive and the long-term reactions (not shown) are insignificant as we had shown in Table 6. From columns (1) and (3), we learn that earnings announcements are more likely to be scheduled on the same day when both firms share few analysts, the subsidiary is relatively large, the parent owns a larger stake in its subsidiary, the parent and subsidiary do not have a common owner, they operate in the same country and share part of a corporate name. Still, the non-significance of Heckman's lambda reveals that failure to condition on strategic announcement timing does not affect the results in the second stage (in the different set-up of the columns (5)-(7), conditioning may have a small effect).

While we test in columns (1) to (4) the simultaneous versus staggered announcement, we also study the robustness of a parent's immediate and delayed reaction to the subsidiary's earnings announcement subject to the possibly endogenous choice of scheduling the subsidiary's announcement first relative to the choice of having the parent announce first. We find that the parent's investor reaction remains unchanged (relative to the findings in Table 7) when controlling for the announcement timing. In column (5), we report the first stage and find that the choice to schedule the subsidiary's announcement first mainly depends on the link characteristics (discussed above).

Overall, we fail to find evidence that strategic timing affects the investor reactions to the parent's and subsidiary's surprise earnings.

[Insert table 3.11 here]

### **3.5.2. Tunneling and parents' expropriation behavior**

We interpret the absence of a subsidiary's investor reaction to the surprise earnings announcement by a parent who announces first, as resulting from investor unawareness of the ownership link. However, an alternative explanation could be fear of tunneling. The rationale is the following: the parent announces a positive earnings surprise, but even if the subsidiary's investors are aware of the ownership link and expect that the positive earnings at the level of the parent result from the subsidiary, they may be skeptical about whether this news is positive for the subsidiary. A positive earnings surprise could for instance reflect that the parent is able to extract earnings from the subsidiary by conducting self-dealing transactions at the expense of the subsidiary's investors. Therefore, positive earnings surprises at the parent level may result from expropriation decisions by the parent, leading

to reduced earnings at the level of the subsidiary. Likewise, an announcement of negative earnings by the parent may indicate that the parent may be enticed to correct these negative earnings by subsequently extracting rents from a well-performing subsidiary. In addition, although we have documented in the section 3.3.2.1 that the correlation between the reported earnings of the parent and subsidiary is very strong, it is still important to check whether our results could be due to ‘tunneling’, which is why we include legal variables as instruments for the potential for expropriation behavior. We use the Djankov et al. (2008) anti-self-dealing index that measures the legal protection of minority investors against self-dealing and expropriation by corporate insiders, and interact the index with the Top Two Quantiles variable. We find that a subsidiary’s investor reaction to the parent’s surprise earnings announcements (that are disclosed prior to those of the subsidiary) are statistically and economically insignificant.<sup>17</sup> We also use an enforcement index by Djankov et al. (2008) that measures the extent to which contracts are enforced in a court of law. We re-estimate our models by including the interaction of the surprise earnings measure with the public enforcement index and do not find any significant relation, which reduces the possibility that our findings are due to potential tunneling.

In case of inverse tunneling, called propping, subsidiaries’ investors should react positively to positive earnings surprises at the parent level, as they could expect, for instance if their own earnings are negative, a capital flow from the parent to the subsidiary (Bae, Cheon and Kang, 2008). As the number of cases of positive earnings surprises by parents with listed subsidiaries incurring low or negative earnings is very small, expected propping is unlikely (or at least may be less anticipated) such that a statistical test is precluded.

### 3.5.3. Internal capital markets

The investor response to surprise earnings announcements may depend on the existence of internal capital markets whereby surpluses in one division are used to fund capital needs in other divisions.<sup>18</sup> For instance, when a subsidiary announces a positive earnings surprise, this may benefit the entire corporate group as the parent could redistribute excess funds to growth-oriented subsidiaries. The parent’s response to the positive earnings surprise of the subsidiary could be stronger if the parent’s performance is poor. Conversely, a negative earnings surprise by the subsidiary may reduce the effectiveness of the internal capital market as redistribution

---

<sup>17</sup> Table is not shown for reasons of parsimony, but is available on request.

<sup>18</sup> For the case of Korean business groups, Almeida, Kim and Kim (2015) show that such chaebols transferred cash from low-growth to high-growth member firms, using cross-firm equity investments. For the case of Chinese business groups, Fan, Jin and Zheng (2016) show that intra-group capital flows are most efficient in the absence of conflicts between controlling and minority shareholders and when external financing constraints are binding.

by the parent is then more difficult. To address this issue, we examine whether investors, observing that their own firm has incurred a negative result (and is a first announcer), react differently to the second announcement (of the affiliated firm). We include in our regressions the variable *Parent Neg. Earnings* that equals one when the parent released negative actual earnings and zero otherwise, and then study the response by parent's investors to a positive earnings surprise disclosed by the subsidiary in a context of growth/value firms (as proxied by Tobin's Q).

Model (1) of Table 12 confirms the positive price reaction by the parent when the subsidiary's earnings surprise is in the top two quantiles of its distribution, which we have shown in Panel A of Table 7. Model (2) reveals that the parent investors' reaction to the subsidiary's announcement is much stronger when the parent had announced negative earnings earlier on, as captured by the interaction term.

We further verify whether a parent's stock price reaction depends on the investment opportunities of parent and subsidiary, as proxied by their market-to-book ratios ( $Q$ ). In model (3), we interact a parent's negative earnings with the subsidiary's *Top Two Quantiles* dummy and the subsidiary's  $Q$ . We find that this triple interaction is negative such that the effect of the interaction term *Sub. Top Two Quantiles x Parent neg. earnings* declines. The positive response of the parent with negative earnings to a positive surprise at the subsidiary level is smaller when that subsidiary has high growth opportunities. The reason could be that the subsidiary is now required to invest more such that fewer funds can be transferred to the parent and this may come at the detriment of the parent's and the other subsidiaries' investment policies.

In Model (4), we run a similar regression but now substitute the subsidiary's  $Q$  by the difference between the subsidiary's and parent's  $Q$  (called  $dQ$ ) and examine whether the interaction of the parent's negative earnings with the subsidiary's positive surprise is affected by the triple interaction term that includes  $dQ$ . We find similar results in that the positive parent investor response to the subsidiary's positive surprise when the parent has negative earnings is smaller when the subsidiary has high and the parent low growth opportunities (high versus low  $Q$ ).

These findings suggest that the magnitude of the stock price reaction to a subsidiary's earnings surprise depends on the parent's earnings and the growth opportunities of the subsidiary relative to those of the parent, which in turn suggests that the existence of internal capital markets could affect price responses.

[Insert table 3.12 here]

### 3.5.4. Confounding events

**Announcement timing effects and PEAD.** When the subsidiary's announcement is scheduled first, parents on average release their earnings 13 calendar days later. The delayed parent's investor reaction to the subsidiary's announcement (table 3.8, Panel A) may not be a post-earnings announcement drift but could be caused by the earnings announcement of the parent itself, which would misdirect our conclusions about the parent's investor ability (not) to perceive ownership connections. In order to address this issue, we first rerun our tests and include a dummy variable equal to one if the parent announces earnings within the 60 trading days after the subsidiary's announcement, which is the period over which we calculated the PEAD. We find that our results about the parent's investor reaction are robust to the inclusion of this contamination dummy (table 3.13, Model (1)).

Second, we rerun the same test on different post-announcement windows with subsamples unaffected by the subsequent parent's announcement. For example, table 3.13, Model (2) tests the parent's investor reaction to the subsidiary's announcement over a 10-day window; the delayed investor reaction is therefore calculated for a period of [2;10] days and the test is performed only on parent-subsidiary annual announcement observations where the parent announces at least 10 trading days subsequent to the subsidiary. Models (3), (4) and (5) report similar tests for the delayed investor reaction calculated over [2;20], [2;30], and [2;40] windows, respectively whereby the parent does not release its earnings within the aforesaid windows. As the sample size significantly declines, we do not restrict the sample to the cases where parent and subsidiary have a common financial year, but add the dummy variable *Same Financial Year*. In addition to the usual control variables (parent's and link characteristics), we estimate the models by including industry pair fixed effects and time (year, month, and day-of-the-week) fixed effects. We find that parent's investor reactions to the subsidiary's announcements essentially remain statistically and economically significant when controlling for announcement contamination. Furthermore, we confirm that a subsidiary's earnings surprise is only gradually priced over time by parent's investors.

[Insert table 3.13 here]

**Multiple parent-subsidiary relations within one corporate network** Our results are not affected by multiple parent-subsidiary relationships within a single corporate network. First, in less than 5% of our parent-subsidiary links, there is a parent with more than one listed subsidiaries and within this group we do not have observations where the subsidiaries make earnings announcements on the same announcement day that is different from the parent's. Second, a robustness test

excluding these cases does not affect the results of tables 6-10.

### **3.5.5. Analysts' updating their forecasts after the first announcement**

Analysts are important agents in capital markets in providing in depth analyses of financial and non-financial information released by firms. Their expertise may be especially valuable in the case of complex firms to help investors understand the links between the different entities in the group. When the subsidiary announces its earnings first, analysts following the parent company should process the subsidiary's earnings announcement, and update their forecast to include the newly released information. Similarly, analysts who follow subsidiaries announcing second should update their forecast subsequent to the parent's announcement. Our definition of earnings surprise is based on the median of the most recent forecasts issued by analysts following the company, which may not reflect the entire set of available information, and the heterogeneity of analyst's forecasts. Analysts who update their forecast in the 7 days (30 days) following the first announcement (and before the second announcement, made by the affiliated firm) represent only 2.4% (5.5%) of all analysts making predictions for the second announcer. These results suggest that in the vast majority of firms there is no update of the analysts' forecasts for the second announcer subsequent to the first announcement, such that financial analysts provide limited assistance to investors to better understand multiple announcements in corporate groups.

## **3.6. Conclusion**

We have examined the impact of ownership complexity in business groups on investor reactions when unanticipated information on earnings by affiliated firms is released. We label the apex company as the parent that is linked to what we call a listed 'subsidiary' and the link is based on direct or indirect equity stakes of at least 20% (which is a minimal threshold for consolidation under IFRS rules).

When the parent releases its earnings prior to those of the subsidiary, the parent's investors react both to the surprise earnings announcement of their own company and to the subsequent announcement by the subsidiary, which implies that the latter announcement still contains additional information not yet priced at the parent's initial announcement. These findings suggest that the network induces *enhanced transparency* for investors who comprehend the ownership links. In contrast, the subsidiary's investors only react to the subsidiary's announcement, ignoring the predictive information released at the parent's level at an earlier stage. This suggests

that the subsidiary's investors may be *inattentive* towards the ownership relation of the subsidiary with its parent company. When the subsidiary is the first to announce its unanticipated earnings, both the subsidiary's and parent's investors immediately incorporate this information in the share prices, but do so only partially as there is a post-earnings announcement drift, which also suggests inattention by part of the shareholders.

The explanation for these findings is that investors do not or at least not clearly observe the internal structure of the corporate group. The inattention is worsened by geographical diversification of affiliated firms and by the use of intermediate investment vehicles between parent and subsidiary (indirect ownership), but cannot be explained by strategic timing of the disclosure of earnings surprises (as the timing of the announcement may be induced by good or bad news), investor inattention induced by a day-of-the-week effect or seasonality, expropriation of a subsidiary's performance by a parent (tunneling), internal capital markets, or synergy-related explanations across industries. The presence of institutional investors do not seem to reduce the opacity of group structures, with the exception of active investors owning shares in both parent and subsidiary companies.

### 3.7. Tables

**Table 3.1. Geographic Breakdown of Connected Firms around the World**

The table reports the geographic dissection of the parent and subsidiary links. Pairs of publicly listed companies and subsidiaries are identified by means of ownership links in Bureau van Dijk's Orbis database for the period 2000 until 2015. The category *Europe* includes Albania, Austria, Belgium, Croatia, Cyprus, Czech Republic, Germany, Denmark, Spain, Finland, France, Ireland, Greece, Hungary, Italy, Lithuania, Luxembourg, The Netherlands, Norway, Poland, Portugal, Romania, Russia, Slovakia, Sweden, Switzerland, and Ukraine. *Asia* comprises China, Hong-Kong, Korea, India, Indonesia, Japan, Malaysia, Singapore, Thailand, Philippine, Pakistan, and Taiwan. The Americas include Canada, Latin America, and the Caribbean islands. The group *Middle East* includes Israel, Turkey, Lebanon, Oman, and Saudi. AU/NZ stands for Australia and New Zealand.

Subsidiary's Region	Parent's Region								Total
	US	GB	Europe	Asia	Africa	Americas	Middle East	AU/NZ	
US	171	2	9	4	0	10	0	0	196
GB	17	16	5	4	4	7	0	1	54
Europe	15	9	235	9	0	9	2	0	279
Asia	14	19	39	1106	1	22	2	3	1206
Africa	5	21	20	1	38	0	1	0	86
Americas	24	10	29	51	3	122	0	3	242
Middle East	2	1	11	3	0	0	61	0	78
AU/NZ	1	1	4	7	3	1	1	22	40
Total	249	79	352	1185	49	171	67	29	2181

**Table 3.2. Descriptive Statistics**

The table reports the distributional characteristics of parents and subsidiaries, and parent-subsidiary links. The sample of link-year observations includes links for which we could match earnings announcements of the parent and the subsidiary in a given year. Earnings announcement dates come from I/B/E/S and cover the period from January 2000 until December 2015. All numbers are in USD. Detailed variable descriptions and the data sources are provided in Appendix A.

	N.	Mean	Sd	P25	Median	P75
<u>Parent Companies' Characteristics</u>						
BHAR(0;1)	7413	0.001	0.050	-0.023	0.000	0.025
BHAR(2;60)	7413	0.000	0.167	-0.091	-0.010	0.077
Surprise	7413	-0.004	0.072	-0.005	0.000	0.005
Market Value (USD million)	7385	11.045	16.825	1.024	3.616	12.296
Q	7378	0.737	0.866	0.266	0.480	0.874
Amihud Illiquidity	6888	0.000	0.002	0.000	0.000	0.000
N. Analysts	7413	13.598	10.540	5.000	12.000	20.000
Has Institutional Owners (%)	7413	56.873	49.529	0.000	100	100
N. Institutional Owners	7413	24.722	35.887	0.000	7.000	37.000
% Institutional Ownership	7413	15.578	21.615	0.000	0.310	26.800
% Mutual Funds	7413	6.925	11.345	0.000	0.850	10.140
% Active Investors	7413	1.453	3.782	0.000	0.000	0.500
% Banks and Insurance	7413	9.380	13.221	0.000	2.120	15.100
% Family Ownership	7413	2.665	8.199	0.000	0.000	2.000
% State Ownership	7413	4.074	11.824	0.000	0.000	0.440
<u>Subsidiaries' Characteristics</u>						
BHAR(0;1)	14353	0.001	0.060	-0.023	0.000	0.024
BHAR(2;60)	14353	-0.000	0.195	-0.103	-0.010	0.083
Surprise	14353	-0.006	0.083	-0.007	0.000	0.005
Market Value (USD million)	14144	3.093	7.657	0.179	0.650	2.303
Q	14115	0.946	1.186	0.302	0.570	1.087
Amihud Illiquidity	7798	0.001	0.008	0.000	0.000	0.000
N. Analysts	14353	7.831	8.289	2.000	5.000	11.000
Has Institutional Owner (%)	14353	50.017	50.002	0.000	100	100
N. Institutional Owners	14353	12.648	22.217	0.000	1.000	17.000
% Institutional Ownership	14353	10.703	17.297	0.000	0.000	16.120
% Mutual Funds	14353	4.794	9.687	0.000	0.000	5.100
% Active Investors	14353	0.970	3.291	0.000	0.000	0.000
% Banks and Insurance	14353	5.793	10.142	0.000	0.000	7.850
% Family Ownership	14353	1.389	4.592	0.000	0.000	0.570
% State Ownership	14353	2.065	7.845	0.000	0.000	0.000
<u>Links Characteristics</u>						
Relative Size (%)	14866	41.258	70.454	3.297	12.385	44.942
dQ = (sub's Q < parent's Q)	14832	0.327	1.160	-0.114	0.117	0.517
Directly Owned (%)	15117	71.926	44.938	0.000	100	100
% Ownership Parent in Sub.	15117	48.602	22.657	30.000	46.000	60.950
Has a Common Analyst (%)	15112	9.271	29.003	0.000	0.000	0.000
Common Name (%)	15117	26.983	44.389	0.000	0.000	100
Same Industry (%)	15117	42.012	49.359	0.000	0.000	100
Same Country (%)	15117	71.853	44.973	0.000	100.000	100
Has a Common Owner (%)	15117	38.228	48.596	0.000	0.000	100
N. Common Owners	15117	4.477	10.172	0.000	0.000	4.000
% Common Ownership	15117	4.395	10.104	0.000	0.000	3.260
% Common Mutual Funds	15117	0.369	2.265	0.000	0.000	0.000
% Common Active Investors	15117	0.052	0.619	0.000	0.000	0.000
% Common Financial Inst.	15117	1.004	3.695	0.000	0.000	0.120

[ht!]

**Table 3.3. Immediate and Delayed Reactions by Parent's and Subsidiary's Investors**

The table presents results on investor reactions to earnings surprises (belonging to the top two quantiles of the distribution). Buy-and-hold-abnormal returns (BHARs) are calculated using the market model over the trading day period  $(-300, -46)$  days and are measured over the  $(0, +1)$  and  $(+2, +60)$  event windows. Columns (1-4) report results of parent investor reactions to the parent's announcement (1-2) and to subsidiary's announcement (3-4). Columns (5-8) report results of subsidiary investor reactions to the subsidiary's announcement (5-6), and to its parent's announcement (7-8). Parent (Subsidiary) controls include the parent's (Subsidiary's) market value, the log of analyst coverage, the Tobin's Q, and the Amihud illiquidity measure. Link controls are the companies' relative market value, percentage of common analysts, percentage of control held by the parent, a direct ownership dummy, dummy variables equal to one if parent and subsidiary are located in the same country, if they share part of the corporate name, and if the parent and its subsidiary have a common institutional investor. All specifications include pair (parent-subsidiary) SIC2 industry fixed effects, and time fixed effects (year, month, and day-of-the-week). Robust t -statistics are reported between brackets. \*, \*\*, and \*\*\* indicate significance at the 10%, 5%, and 1% level, respectively.

	Parent investors' reaction				Subsidiary investors' reaction			
	P' announcement		S' announcement		S' announcement		P' announcement	
	[0;1]	[2;60]	[0;1]	[2;60]	[0;1]	[2;60]	[0;1]	[2;60]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Announcer Top Two Quantiles	0.0151*** (7.191)	0.0079 (1.102)	0.0024** (2.301)	0.0080* (1.675)	0.0218*** (8.074)	0.0256*** (3.171)	0.0013 (0.717)	0.0109 (1.433)
Same Day	0.0027 (0.825)	-0.0133 (-1.234)	0.0003 (0.118)	0.0084 (0.951)	-0.0001 (-0.034)	0.0036 (0.281)	-0.0020 (-0.589)	-0.0024 (-0.188)
Subsidiary First	-0.0021 (-0.824)	-0.0047 (-0.569)	-0.0000 (-0.019)	0.0125* (1.957)	0.0050 (1.588)	0.0148 (1.579)	0.0013 (0.585)	-0.0201** (-2.156)
Same Fiscal Year	0.0045 (1.179)	-0.0145 (-1.186)	-0.0004 (-0.222)	0.0098 (1.161)	-0.0008 (-0.162)	-0.0025 (-0.223)	0.0066** (1.998)	0.0186 (1.387)
Amihud Illiquidity	2.1465 (1.296)	1.1290 (0.532)	-0.1748 (-0.377)	4.5949* (1.796)	0.1518 (1.006)	-0.7623* (-1.652)	-0.0880 (-0.264)	-1.3678*** (-3.266)
Market Value	0.0019** (1.970)	0.0115*** (3.858)	-0.0002 (-0.378)	0.0076*** (3.074)	0.0032** (2.247)	0.0236*** (7.051)	0.0001 (0.174)	0.0169*** (4.218)
Q	-0.0015 (-0.738)	0.0106* (1.670)	0.0003 (0.379)	0.0124*** (3.539)	0.0025 (1.395)	0.0090* (1.903)	0.0034 (1.331)	0.0037 (0.597)
N. Analysts	-0.0015 (-0.934)	-0.0202*** (-3.607)	-0.0002 (-0.270)	-0.0096** (-2.447)	-0.0062** (-2.482)	-0.0319*** (-4.951)	0.0014 (0.882)	-0.0089 (-1.251)
Relative Market Value			0.0002 (0.159)	0.0072 (1.089)			-0.0002 (-0.152)	-0.0150** (-2.389)
Has a Common Owner			0.0002 (0.108)	0.0034 (0.493)			-0.0021 (-0.731)	0.0028 (0.241)
% Common Analysts			0.0211* (1.699)	0.0402 (0.774)			-0.0170 (-1.240)	-0.1010 (-1.587)
% Ownership Parent in Sub.			0.0000 (0.437)	0.0000 (0.173)			-0.0000 (-0.844)	-0.0000 (-0.031)
Common Name			0.0015 (1.011)	0.0060 (0.839)			-0.0025 (-0.938)	-0.0093 (-0.881)
Directly Owned			-0.0011 (-0.659)	-0.0095 (-1.254)			0.0029 (1.029)	-0.0117 (-0.961)
Same Country			-0.0010 (-0.584)	0.0112 (1.401)			0.0035 (1.271)	0.0061 (0.535)
Subsidiary Industry FE	No	No	No	No	Yes	Yes	No	No
Parent Industry FE	Yes	Yes	No	No	No	No	No	No
Pair Industry FE	No	Yes	Yes	Yes	No	No	Yes	Yes
Year+Month+Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.038	0.105	0.014	0.119	0.034	0.140	0.073	0.167
N	2682	2682	5331	5331	3031	3031	3023	3023

**Table 3.4. Calendar Days Distance between Parent and Subsidiary Announcements**

The table reports the distribution of time distance, in calendar days, between a parent's earnings announcement and its subsidiary in a given year. The sample is partitioned into (i) link-year observations where the parent and its subsidiary close their financial year on the same date and (ii) link-year observations where the parent and its subsidiaries do not. The samples are then further partitioned into three subsamples: (i) the parent and subsidiary make their earnings announcement on the same day, (ii) the subsidiary releases its earnings information first, and (iii) the parent releases its earnings first.

	N. Obs.	Mean	Sd	P25	Median	P75
Parent and Subsidiary (with a link) with the Same Financial Year						
Same Day	1841	0	0	0	0	0
Subsidiary First	5523	12.9	17.9	2	7	15
Parent First	4337	22.6	28.5	6	14	28
Parent and Subsidiary (with a link) with a different Financial Year						
Same Day	4	0	0	0	0	0
Subsidiary First	1282	101.9	70.6	50	86	141
Parent First	990	142.1	80.7	85	118.5	198

**Table 3.5. Does the Information Released by the First Announcer Matter?**

The table presents OLS estimates and verifies whether information released by the first announcer is correlated to information released by the second announcement. Columns (1-2) correspond to the situations where the subsidiary announces first. Column 1 reports results from regression of the parent's actual earnings on the subsidiary's actual earnings. Column 2 reports results from regression of the parent's *Top Two Quantiles* on the subsidiary's *Top Two Quantiles*. Columns (3-4) correspond to the situations where the parent is the first announcer. Column 3 reports results from regression of the subsidiary's actual earnings on the parent's actual earnings. Column 4 reports results from regression of the subsidiary's *Top Two Quantiles* on the parent's *Top Two Quantiles*. Parents' *Top Two Quantiles* and subsidiary's *Top Two Quantiles* are calculated with respect to the parent's actual earnings distribution and subsidiary's actual earnings distribution, respectively. All models include link controls, pair (parent-subsidiary) industry SIC-2 fixed effects, and time (year, month and day-of-the-week) fixed effects. Link controls are the companies' relative market value, percentage of common analysts, percentage of control held by the parent, a direct ownership dummy, dummy variables equal to one if parent and subsidiary are located in the same country, if they share part of the corporate name, and if the parent and its subsidiary have a common institutional investor. Robust t-statistics are reported between brackets. \*, \*\*, and \*\*\* indicate significance at the 10%, 5%, and 1% level, respectively.

	Subsidiary Announces First		Parent Announces First	
	Parent Actual Earnings (1)	Parent Top Two Quantiles (2)	Subsidiary Actual Earnings (3)	Subsidiary Top Two Quantiles (4)
Subsidiary Actual Earnings	0.3619*** (3.401)			
Subsidiary Top Two Quantiles		0.1919*** (5.177)		
Parent Actual Earnings			0.0501*** (3.095)	
Parent Top Two Quantiles				0.2254*** (4.645)
Relative Market Value	-0.1818 (-0.870)	0.0119 (0.468)	0.5092*** (3.276)	0.0679 (1.637)
% Common Analysts	3.4475** (2.295)	0.2887 (0.907)	-3.6097*** (-2.850)	-0.7197* (-1.956)
Has a Common Owner	0.5529*** (2.586)	0.0863* (1.681)	0.6935** (2.445)	-0.0284 (-0.450)
% Ownership Parent in Sub.	-0.0140** (-2.290)	0.0005 (0.360)	-0.0108** (-2.184)	-0.0012 (-0.999)
Common Name	-0.1923 (-1.004)	-0.0013 (-0.027)	0.5279** (1.988)	-0.0009 (-0.014)
Directly Owned	-0.9821*** (-2.933)	0.0395 (0.641)	-1.6978*** (-6.419)	0.0452 (0.616)
Same Country	0.7407** (2.368)	0.1079* (1.704)	-0.6858** (-2.559)	-0.0324 (-0.454)
Pair Industry FE	Yes	Yes	Yes	Yes
Year+Month+Day FE	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.193	0.183	0.359	0.099
N	5740	954	4459	673

**Table 3.6. Parent and Subsidiary Announce on Same Day**

The table presents results from regressions of the parent and subsidiary investor reactions to announcements of earnings surprises (belonging to the top two quantiles of their distribution), for the cases where the parent and its subsidiary close their financial year on the same day and their earnings announcements take place on the same day. Buy-and-hold-abnormal returns (BHARs) are calculated using the market model over the trading day period (-300,-46 days) and are measured over the (0,+1) and (+2,+60) event windows. Columns (1-4) report parent's investor reactions to the earnings announcements and columns (5-8) report subsidiary's investor reactions to the earnings announcements. Specifications in the even-numbered columns report results with pair (parent-subsidiary) industry SIC2 fixed-effects. Parent (subsidiary) controls are the parent's (subsidiary's) market value, the log of analyst coverage, the Tobin's Q, and the Amihud illiquidity measure. All models include time (year, month, and day-of-the-week) fixed effects. Robust t-statistics are reported between brackets. \*, \*\*, and \*\*\* indicate significance at the 10%, 5%, and 1% level, respectively.

	Parent investors' reaction				Subsidiary investors' reaction			
	P' announcement		S' announcement		S' announcement		P' announcement	
	[0;1]	[2;60]	[0;1]	[2;60]	[0;1]	[2;60]	[0;1]	[2;60]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Top Two Quantiles	0.0163*** (3.453)	0.0166*** (2.607)	0.0058 (0.358)	-0.0049 (-0.228)	0.0224*** (3.024)	0.0277*** (3.810)	-0.0302 (-1.228)	-0.0573* (-1.706)
Amihud Illiquidity		0.5996 (1.380)		3.1066** (2.285)		0.8200* (1.855)		-1.3680* (-1.833)
Market Value		0.0037 (1.237)		0.0088 (0.861)		0.0018 (0.314)		0.0332 (1.467)
Q		0.0033 (1.054)		-0.0092 (-1.052)		-0.0061 (-0.895)		0.0117 (0.539)
N. Analysts		-0.0023 (-0.456)		-0.0250 (-1.383)		0.0075 (0.780)		-0.0913*** (-2.741)
Relative Market Value		0.0038 (0.781)		-0.0129 (-0.737)		-0.0108 (-1.155)		0.0247 (0.756)
Common ownership		0.0020 (0.210)		0.0038 (0.105)		0.0073 (0.533)		0.0055 (0.101)
% Common Analysts		-0.0825 (-1.381)		-0.0093 (-0.047)		-0.0603 (-0.877)		-0.1158 (-0.391)
% Ownership Parent in Sub.		-0.0003 (-1.200)		0.0003 (0.339)		0.0004 (1.018)		0.0011 (1.065)
Common Name		0.0089 (0.911)		0.0043 (0.157)		-0.0115 (-0.987)		0.0274 (0.653)
Directly Owned		-0.0174 (-1.607)		-0.0426 (-1.087)		-0.0093 (-0.633)		0.0262 (0.540)
Same Country		-0.0068 (-0.620)		0.0584 (1.401)		0.0008 (0.060)		-0.1536** (-2.000)
Pair Industry FE	No	Yes	No	Yes	No	Yes	No	Yes
Year+Month+Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.009	0.010	0.052	0.032	-0.006	0.066	0.119	0.099
N	470	385	470	385	360	291	360	291

**Table 3.7. Parent Announces First**

The table presents results from regressions of investors' reactions to announcements of earnings surprises (belonging to the top two quantiles of their distribution) for the cases where the parent's and its subsidiary's financial years coincide and the parent releases its earnings prior to the subsidiary. Panel A reports the parent investor reactions to the announcement by the parent (columns 1-4), and to the subsidiary' announcement (columns 5-8). Panel B reports the subsidiary investor reaction to the parent's announcement (columns 1-4), and to the announcement of the subsidiary (columns 5-8). All specifications include year, month, and day-of-the-week fixed-effects. The even columns include firm controls and firm SIC-2 industry fixed effects for the models of investor reactions to their own company's announcement. If models test investors' reactions to the affiliated firm's announcement, they include pair (parent-subsidiary) industry SIC-2 fixed effects, and parent-subsidiary link controls in addition to firm controls. Parent (subsidiary) controls include the parent's (subsidiary's) market value, the log of analyst coverage, market-to-book ratio, and the Amihud illiquidity measure. Link controls comprise subsidiary-parent relative market value, percentage of control held by the parent, percentage of common analysts, a direct ownership dummy, dummy variables equal to one if parent and subsidiary are located in the same country, if they share part of a corporate name, and if common institutional investors own a stake in the parent and the subsidiary, respectively. Robust t-stats are reported between brackets. \*, \*\*, and \*\*\* indicate significance at the 10%, 5%, and 1% level, respectively.

	Parent investors' reaction				Subsidiary investors' reaction			
	P' announcement		S' announcement		S' announcement		P' announcement	
	[0;1]	[2;60]	[0;1]	[2;60]	[0;1]	[2;60]	[0;1]	[2;60]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Announcer Top Two Quantiles	0.0117***	0.0106**	-0.0097	-0.0174	0.0034**	0.0049***	-0.0074	-0.0099
	(2.813)	(2.551)	(-0.697)	(-1.211)	(2.250)	(2.864)	(-1.088)	(-1.192)
Parent Amihud Illiquidity		4.3948*		0.0375		-0.7923***		9.7426***
		(1.836)		(0.027)		(-3.624)		(12.704)
Parent Market Value		-0.0005		0.0089		-0.0027**		0.0136***
		(-0.267)		(1.420)		(-2.517)		(2.936)
Parent Q		0.0003		0.0700***		0.0045**		0.0193**
		(0.067)		(3.255)		(2.145)		(1.982)
Parent N. Analysts		-0.0016		-0.0054		0.0026		-0.0146
		(-0.433)		(-0.508)		(1.144)		(-1.525)
Link Controls	No	No	No	No	No	Yes	No	Yes
Parent Industry FE	No	Yes	No	Yes	No	No	No	No
Pair Industry FE	No	No	No	No	No	Yes	No	Yes
Year+Month+Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.005	0.080	0.112	0.137	0.007	0.022	0.093	0.114
N	679	668	679	668	1866	1762	1866	1762

Parent Announces First (Cont'd)

Panel B: Subsidiary's Investor Reactions

	Subsidiary investors' reactions							
	Parent's announcement				Subsidiary's announcement			
	BHAR [0;1]	BHAR [0;1]	BHAR [2;60]	BHAR [2;60]	BHAR [0;1]	BHAR [0;1]	BHAR [2;60]	BHAR [2;60]
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Announcer Top Two Quantiles	0.0003	-0.0023	0.0028	0.0014	0.0253***	0.0243***	0.0562***	0.0578***
	(0.096)	(-0.609)	(0.204)	(0.089)	(7.318)	(6.168)	(4.440)	(4.168)
Subsidiary Amihud Illiquidity		-0.7034		-1.5680***		-0.1369		0.0714
		(-1.084)		(-2.963)		(-0.817)		(0.108)
Subsidiary Market Value		0.0021		0.0287***		0.0042**		0.0180***
		(1.061)		(2.928)		(1.986)		(2.981)
Subsidiary Q		0.0012		0.0076		0.0032		0.0156
		(0.713)		(0.862)		(1.100)		(1.641)
Subsidiary N. Analysts		-0.0029		-0.0160		-0.0054*		-0.0357***
		(-0.791)		(-1.200)		(-1.709)		(-3.084)
Link Controls	No	Yes	No	Yes	No	No	No	No
Subsidiary Industry FE	No	No	No	No	No	Yes	No	Yes
Pair Industry FE	No	Yes	No	Yes	No	No	No	No
Year+Month+Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	-0.001	-0.028	0.159	0.196	0.053	0.063	0.123	0.151
N	871	798	871	798	1133	1129	1133	1129

**Table 3.8. Subsidiary Announces First**

The table presents results from regressions of investors' reactions to earnings announcement of earnings surprises (belonging to the top two quantiles of their distribution), for the cases where parent and subsidiary close their financial year on the same date and where the subsidiary releases its earnings first. Panel A reports the parent's reactions to the subsidiary's announcement (columns 1-4), and to the announcement of the parent which takes place after the subsidiary's (columns 5-8). Panel B reports the subsidiary's investor reactions to the subsidiary's announcement (columns 1-4), and to the parent's announcement (columns 5-8). All specifications report results with year, month, and day-of-the-week fixed-effects. The even-numbered columns include firm controls and firm SIC-2 industry fixed effects, if they concern investor's reactions to their own company's announcement. If models test investors' reaction to the affiliated company's announcement, they include pair (parent-subsidiary) industry SIC-2 fixed effects and link (parent-subsidiary) controls in addition to firms' controls. Parent (subsidiary) controls comprise the parent's (subsidiary's) market value, log of analyst coverage, the Tobin's Q, and the Amihud illiquidity measure. Link controls comprise subsidiary-parent relative value, percentage of control held by the parent, percentage of common analysts, a directly ownership dummy, dummy variables equal to one if parent and subsidiary are located in the same country, share (part of) a corporate name, and if they share a common institutional owner, respectively. Robust t-stats are reported between brackets. \*, \*\*, and \*\*\* indicate significance at the 10%, 5%, and 1% level, respectively.

	Panel A. Parent's Investor Reactions							
	Subsidiary's announcement				Parent investors' reactions			
	BHAR [0;1]		BHAR [2;60]		Parent's announcement		BHAR [0;1]	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Announcer Top Two Quantiles	0.0036** (2.392)	0.0031* (1.677)	0.0287*** (3.689)	0.0277*** (3.205)	0.0124*** (4.352)	0.0108*** (3.832)	0.0196* (1.835)	0.0186* (1.676)
Parent Amihud Illiquidity		0.2668 (0.316)		19.7968 (1.111)		-1.6293 (-1.329)		11.3764 (0.701)
Parent Market Value		-0.0007 (-0.629)		0.0095* (1.901)		0.0012 (0.864)		0.0175*** (3.252)
Parent Q		-0.0005 (-0.355)		0.0148* (1.704)		0.0020 (0.777)		0.0017 (0.144)
Parent N. Analysts		-0.0000 (-0.015)		-0.0178** (-2.507)		-0.0014 (-0.603)		-0.0282*** (-3.195)
Link Controls	No	Yes	No	Yes	No	No	No	No
Parent Industry FE	No	No	No	No	No	Yes	No	Yes
Pair Industry FE	No	Yes	No	Yes	No	No	No	No
Year+Month+Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.014	0.033	0.123	0.175	0.010	0.013	0.144	0.148
N	2083	1960	2083	1960	1147	1139	1147	1139

**Subsidiary Announces First (Cont'd)**

Panel B. Subsidiary's Investor Reactions

	Subsidiary investors' reactions							
	Subsidiary's announcement				Parent's announcement			
	BHAR [0;1]		BHAR [2;60]		BHAR [0;1]		BHAR [2;60]	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Announcer Top Two Quantiles	0.0253***	0.0243***	0.0562***	0.0578***	0.0014	-0.0005	0.0317***	0.0301**
	(7.318)	(6.168)	(4.440)	(4.168)	(0.478)	(-0.165)	(2.799)	(2.273)
Subsidiary Amihud Illiquidity		-0.1369		0.0714		0.9656		-0.7069
		(-0.817)		(0.108)		(1.243)		(-1.125)
Subsidiary Market Value		0.0042**		0.0180***		0.0014		0.0152*
		(1.986)		(2.981)		(0.854)		(1.681)
Subsidiary Q		0.0032		0.0156		0.0042**		0.0112*
		(1.100)		(1.641)		(2.349)		(1.743)
Subsidiary N. Analysts		-0.0054*		-0.0357***		-0.0001		-0.0126
		(-1.709)		(-3.084)		(-0.047)		(-0.851)
Link Controls	No	No	No	No	No	Yes	No	Yes
Subsidiary Industry FE	No	Yes	No	Yes	No	No	No	No
Pair Industry FE	No	No	No	No	No	Yes	No	Yes
Year+Month+Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.053	0.063	0.123	0.151	0.007	0.177	0.107	0.195
N	1133	1129	1133	1129	1409	1334	1409	1334

**Table 3.9. Channels of Parent Investor Unawareness**

The table presents results from regressions of parent investor reactions to subsidiary's earnings announcement when parent and subsidiary close their financial year on the same date and the subsidiary releases its earnings first. Panel A reports immediate and delayed reactions to subsidiary's announcement when parent and subsidiary are located in the same country (columns 1-2) or in different countries (col. 3-4). Panel B reports immediate and delayed reactions to subsidiary's announcement when the subsidiary is directly owned by the parent (col. 1-2) and is indirectly controlled by the parent (columns 3-4). Panel C reports the parent reaction to subsidiary's announcement when parent and subsidiary share part of a corporate name (columns 1-2), and do not (columns 3-4). Panel D reports the parent reaction to the subsidiary's announcement when the parent controls less than 50% (columns 1-2), and more than 50% of the subsidiary (columns 3-4). Specifications in panels A, B and C include year, month, day-of-the-week fixed-effects, pair (parent-subsidiary) industry SIC-2 fixed effects, parent and link controls. Note that, specifications in panel D include time, parents' controls, and link controls and do not include pair industry fixed effects. Parent controls comprise the parent's market value, log of analyst coverage, the Tobin's Q, and the Amihud illiquidity measure. Link controls include subsidiary-parent relative market value, percentage of control held by the parent, percentage of common analysts, a direct ownership dummy, dummy variables equal to one if both firms located in the same country, operate in the same SIC-2 industry, share part of a corporate name, and share a common institutional owner, respectively. Robust t-stats are reported between brackets. \*, \*\*, and \*\*\* indicate significance at the 10%, 5%, and 1% level, respectively.

	Parent investor reactions to Subsidiary's announcement (Subsidiary announces first)			
	(1)	(2)	(3)	(4)
	Panel A	Same Country		Different Country
	[0;1]	[2;60]	[0;1]	[2;60]
Subsidiary Top Two Quantiles	0.0044** (1.977)	0.0276*** (2.661)	0.0003 (0.095)	0.0291* (1.814)
N	1433	1433	514	514
Panel B	Directly Owned		Indirectly Owned	
	[0;1]	[2;60]	[0;1]	[2;60]
Subsidiary Top Two Quantiles	0.0042** (1.983)	0.0263** (2.533)	0.0029 (0.865)	0.0306* (1.873)
N	1421	1420	518	518
Panel C	Common Name		Different Name	
	[0;1]	[2;60]	[0;1]	[2;60]
Subsidiary Top Two Quantiles	0.0047* (1.653)	0.0308** (2.283)	0.0030* (1.650)	0.0219** (2.316)
N	699	699	1384	1384
Panel D	Ownership < 50%		Ownership > 50%	
	[0;1]	[2;60]	[0;1]	[2;60]
Subsidiary Top Two Quantiles	0.0035* (1.715)	0.0230** (2.168)	0.0035 (1.525)	0.0300*** (2.692)
N	1050	1050	1031	1031
Parent Controls	Yes	Yes	Yes	Yes
Link Controls	Yes	Yes	Yes	Yes
Pair Industry FE	Yes	Yes	Yes	Yes
Year+Month+Day FE	Yes	Yes	Yes	Yes

**Table 3.10. Investor Sophistication**

The table presents results from regressions of parent investor reactions to subsidiary's earnings announcement, for the cases where parent and subsidiary close their financial year on the same date and where the subsidiary releases earnings first. Panel A studies the impact of the type of parents' institutional minority owners on the reaction to the subsidiaries' announcements. Columns (1-2) include the cumulative ownership percentages owned by types of institutional minority owners. Panel B studies the effect of common institutional minority ownership in the parent and in the subsidiary on the reaction of parent investors to the subsidiary's announcement. Columns (1-2) present results from regressions including cumulative ownership percentages owned by types of institutional minority owners that own a stake both in the parent and its subsidiary. We distinguish three types of institutional owners: mutual funds, active investors (PE, VC, HF), and banks and insurance companies (including other financial companies). All specifications report results with year, month, day-of-the-week fixed-effects, pair (parent-subsidiary) industry SIC-2 fixed effects, parents' controls, and link controls. Parent controls comprise the parent's market value, log of analyst coverage, the Tobin's Q, and the Amihud illiquidity measure. Link controls include subsidiary-parent relative market value, percentage of control held by the parent, percentage of common analysts, a directly ownership dummy, dummy variables equal to one if parent and subsidiary are located in the same country, operate in the same SIC-2 industry, share (part of) a corporate name, and if they share a common institutional owner, respectively. Robust t-stats are reported between brackets. \*, \*\*, and \*\*\* indicate significance at the 10%, 5%, and 1% level, respectively.

Panel A. Institutional Owners in the Parent Company		
	BHAR[0;1]	BHAR[2;60]
	(1)	(2)
Subsidiary Top Two Quantiles	0.0047** (2.403)	0.0181* (1.907)
% Mutual Funds	0.0005*** (3.351)	0.0002 (0.196)
Top Two Quantiles x % Mutual Funds	-0.0006*** (-2.992)	0.0007 (0.643)
% Active investors	-0.0002 (-0.496)	-0.0011 (-0.712)
Top Two Quantiles x % Active investors	0.0009 (1.437)	0.0021 (0.937)
% Banks and Insurance	0.0001 (0.798)	-0.0012* (-1.765)
Top Two Quantiles x % Banks and Insurance	0.0001 (0.794)	0.0002 (0.185)
% Family Ownership	-0.0000 (-0.295)	-0.0006 (-1.176)
% State Ownership	0.0000 (0.119)	-0.0003 (-1.088)
Parent Controls	Yes	Yes
Link Controls	Yes	Yes
Pair Industry FE	No	No
Year+Month+Day FE	Yes	Yes
R <sup>2</sup>	0.021	0.147
N	2083	2083

**Investor Sophistication (Cont'd)**

Panel B. Common Institutional Owners

	BHAR[0;1] (1)	BHAR[2;60] (2)
Subsidiary Top Two Quantiles	0.003** (2.128)	0.027*** (3.346)
% Common Mutual Funds	-0.000 (-0.463)	0.003 (0.687)
Top Two Quantiles x % Common Mutual Funds	0.000 (0.318)	0.002 (0.400)
% Common Active Owners	0.005*** (2.726)	-0.002 (-0.159)
Top Two Quantiles x % Common Active Owners	0.007** (2.458)	0.005 (0.332)
% Common Banks and Insurance	0.000 (1.005)	0.001 (0.567)
Top Two Quantiles x % Common Banks and Insurance	-0.000 (-1.050)	-0.002 (-1.299)
% Family Ownership	-0.000 (-0.143)	-0.001 (-1.177)
% State Ownership	-0.000 (-0.269)	-0.000 (-0.966)
Parent Controls	Yes	Yes
Link Controls	Yes	Yes
Pair Industry FE	No	No
Year+Month+Day FE	Yes	Yes
$R^2$	0.014	0.145
N	2083	2083

**Table 3.11. Strategic Announcement Timing and Investor Reactions**

The table tests the strategic disclosure hypothesis: The models in columns 1-2 (3-4) report results of parent's (subsidiary's) investor reactions to the earnings announcements of the subsidiary (announcing first) while endogenizing the parent's and subsidiary's decision about when to announce (on same day versus on different days). First stage specifications include parent's and subsidiary's earnings. Columns (1) and (3) report first stage results of parent's and subsidiary's announcement timing, respectively. The models in columns (5-7) estimates parent's investor immediate (column (6)) and longer-term (column (7)) reactions to the subsidiary's announcement conditional on the subsidiary announcing first (relative to the parent announcing first). First stage specifications include pair industry and time fixed effects. Second stage specifications include pair industry fixed effects, year, month and day-of-the-week fixed effects. Robust t-tats are between brackets. \*, \*\*, and \*\*\* indicate significance at 10, 5, and 1%, respectively.

Investors Heckman stage Dependent variable	Parent's reaction		Subsidiary's reaction		Parent's reaction		
	1st Stage Same Day (1)	2nd Stage [0;1] (2)	1st Stage Same Day (3)	2nd Stage [0;1] (4)	1st Stage Sub. First (5)	2nd Stage [0;1] (6)	2nd Stage [2;60] (7)
Top Two Quantiles		0.0176*** (3.4732)		0.0315*** (4.2155)		0.0031* (1.9104)	0.0326*** (3.9989)
Market Value		0.0051*** (2.7488)		0.0018 (0.4574)		0.0007 (1.0712)	0.0185*** (5.4806)
Amihud Illiquidity		0.0005 (0.8558)		0.0000 (0.2812)		0.0002 (0.1785)	0.0153*** (3.6150)
N. Analysts		-0.0036 (-0.9408)		0.0027 (0.3683)		-0.0025* (-1.8566)	-0.0333*** (-4.9880)
% Common Analysts	-1.8803*** (-3.8094)	-0.0575 (-0.6388)	0.1357 (0.3704)	0.0314 (0.5191)	-0.6615** (-2.4041)	0.0013 (0.0872)	0.0266 (0.3315)
Relative Market Value	0.2042*** (6.6806)	-0.0089 (-1.0846)	0.1413*** (3.9339)	-0.0141 (-1.0423)	0.2463*** (8.2233)	0.0049* (1.7143)	-0.0245 (-1.5588)
Has a Common Owner	-0.1810*** (-2.7051)	0.0063 (0.6363)	-0.0373 (-0.5806)	0.0044 (0.4301)	-0.1485*** (-3.3959)	-0.0031 (-1.1958)	0.0201 (1.4024)
% Ownership P in Sub.	0.0056*** (4.2657)	-0.0003 (-1.1568)	0.0069*** (5.6393)	0.0001 (0.2383)	0.0048*** (5.5693)	0.0001* (1.8609)	-0.0009** (-2.5368)
Directly Owned	0.0993 (1.3240)	-0.0118 (-1.4103)	0.1108 (1.5153)	-0.0095 (-0.6051)	0.1056** (2.2716)	-0.0022 (-0.8837)	-0.0456*** (-3.3808)
Same Country	0.3388*** (4.2868)	-0.0179 (-1.1698)	0.0436 (0.5889)	0.0166 (1.4642)	0.1744*** (3.6763)	0.0001 (0.0416)	0.0269* (1.6652)
Common Name	0.3635*** (5.9483)	-0.0085 (-0.5678)	0.1969*** (3.4250)	-0.0088 (-0.4584)	0.3944*** (9.5202)	0.0042 (0.8968)	-0.0388 (-1.5085)
Parent Surprise	-0.5439* (-1.7555)		0.4958 (1.1722)		-0.4477** (-1.9735)		
Subsidiary Surprise	-0.2647 (-0.6733)		-0.3384 (-1.0484)		-0.2986* (-1.6509)		
Lambda		-0.054 (-1.15)		-0.018 (-0.19)		-0.166* (-1.87)	
N	5,294	5,294	10,286	10,286	6,699	6,699	6,699
Pair Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	No	Yes	No	No	Yes	Yes
Day FE	Yes	No	Yes	No	No	Yes	Yes

**Table 3.12. Parent's Subsidizing Behavior**

The table investigates possible effects of internal capital markets by regressing the parents' investor reactions on the subsidiaries' earnings announcements, for the case where parent and subsidiary close their financial year on the same date and where the parent releases earnings first. *Parent Neg. Earnings* is equal to one if the parent announces negative earnings, and zero otherwise. *Q* is the subsidiary's market-to-book ratio. *dQ* is the difference between subsidiary's and parent's market-to-book ratios. All models include pair industry fixed effects, year, month, day-of-the-week fixed effects, parent controls (parent's market value, log of analyst coverage, and the Amihud illiquidity measure), and link controls (relative market value, percentage of control held by parent, percentage of common analysts, a direct ownership dummy, dummy variables equal to one if parent and subsidiary are located in the same country, if they share part of a corporate name, and if they share a common institutional investors, respectively). Robust t-stats are reported between brackets. \*, \*\*, and \*\*\* indicate significance at the 10%, 5%, and 1% level, respectively.

Parent investors' response:	BHAR[0;1] (1)	BHAR[0;1] (2)	BHAR[0;1] (3)	BHAR[0;1] (4)
Subsidiary Top Two Quantiles	0.0045*** (2.610)	0.0031* (1.673)	0.0018 (0.776)	0.0020 (1.042)
Parent Neg. Earnings	-0.0012 (-0.458)	-0.0034 (-1.153)	-0.0033 (-0.998)	-0.0038 (-1.285)
Q (Subsidiary)	0.0028*** (2.621)	0.0028*** (2.619)	-0.0013 (-0.696)	
Sub. Top Two Quantiles × Parent Neg. Earnings		0.0087* (1.695)	0.0126** (2.048)	0.0098* (1.899)
Sub. Top Two Quantiles × Q (Subsidiary)			-0.0036 (-1.576)	
Parent Neg. Earnings × Q (Subsidiary)			-0.0004 (-0.181)	
Sub. Top Two × Parent Neg. Earnings × Q (S)			-0.0057* (-1.775)	
dQ (Subsidiary. Q-Parent Q)				-0.0015 (-0.940)
Sub. Top Two Quantiles × dQ				0.0057** (2.498)
Parent Neg. Earnings × dQ				-0.0047** (-2.096)
Sub. Top Two × Parent Neg. Earnings × dQ				-0.0090*** (-2.845)
Q (Parent)	0.0049** (2.300)	0.0049** (2.283)	0.0051** (2.396)	
Link Controls	Yes	Yes	Yes	Yes
Pair Industry FE	Yes	Yes	Yes	Yes
Year+Month+Day FE	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.027	0.028	0.028	0.022
N	1731	1731	1731	1731

**Table 3.13. Confounding Events**  
**Parent Investor Reaction to Subsidiary's Announcements**

The table presents results from regressions of parent investor reactions to the subsidiary's earnings announcements of earnings surprises (belonging to the top two quantiles of their distribution) for the case where the subsidiary releases its earnings information first. Observations with parent and subsidiary closing their financial year at different dates are also included in this sample. Column (1) reports parent's investors delayed reaction to the subsidiary's announcement over [+2;+60] trading days after the announcement. A *contaminated window* variable is included and takes the value one if the parent announces within a period of 60 trading days after the subsidiary's announcement (over which the delayed reaction is calculated). Columns (2-5) report delayed parent investor reactions to the subsidiary's announcement, calculated for different event windows prior to the parent's announcement - observations with parent's announcements occurring within the event window are here excluded. All specifications include a dummy variable *same financial year* that is equal to one if parent and subsidiary close their financial year on the same date. The specifications include parent controls (market value, log of analyst coverage, Tobin's Q, and the Amihud illiquidity measure), time-varying link controls (relative market value, percentage of common analysts, and a dummy common institutional outside investor), pair (parent-subsidiary) fixed effects and time (year, month, day-of-the-week) fixed effects. Robust t-stats are reported between brackets. \*, \*\*, and \*\*\* indicate significance at the 10%, 5%, and 1% level, respectively.

Time window	BHAR[2;60] (1)	BHAR[2;10] (2)	BHAR[2;20] (3)	BHAR[2;30] (4)	BHAR[2;40] (5)
Subsidiary Top Two Quantiles	0.0232*** (3.114)	0.0093 (1.328)	0.0081 (0.997)	0.0233* (1.834)	0.0337** (2.075)
Same Financial Year	0.0019 (1.136)	-0.0627** (-2.045)	-0.0238 (-1.348)	-0.0127 (-0.349)	-0.0307 (-0.623)
Contaminated Window	0.0043 (0.454)				
Parent Controls	Yes	Yes	Yes	Yes	Yes
Link Controls	Yes	Yes	Yes	Yes	Yes
Pair Industry FE	Yes	Yes	Yes	Yes	Yes
Year+Month+Day FE	Yes	Yes	Yes	Yes	Yes
$R^2$	0.152	0.056	0.066	0.079	0.068
N	2444	335	638	476	413

# Bibliography

- Abarbanell, Jeffery S, and Victor L Bernard.** 1992. “Tests of analysts’ overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior.” *Journal of Finance*, 47(3): 1181–1207.
- Abowd, John M, and Francis Kramarz.** 2003. “The costs of hiring and separations.” *Labour Economics*, 10(5): 499–530.
- Abowd, John M, Francis Kramarz, and David N Margolis.** 1999. “High wage workers and high wage firms.” *Econometrica*, 67(2): 251–333.
- Adams, Renée B, and Daniel Ferreira.** 2009. “Women in the boardroom and their impact on governance and performance.” *Journal of Financial Economics*, 94(2): 291–309.
- Adams, Renée B, and Patricia Funk.** 2012. “Beyond the glass ceiling: Does gender matter?” *Management Science*, 58(2): 219–235.
- Adams, Renée B, and Vanitha Ragunathan.** 2017. “Lehman sisters.” *Working Paper*.
- Aghion, Philippe, Ufuk Akcigit, Matthieu Lequien, and Stefanie Stantcheva.** 2017. “Do entrepreneurship and self-employment respond to simpler fiscal incentives? Evidence from France.” *Working Paper*.
- Agrawal, Ashwini K, and David A Matsa.** 2013. “Labor unemployment risk and corporate financing decisions.” *Journal of Financial Economics*, 108(2): 449–470.
- Ahern, Kenneth R, and Jarrad Harford.** 2014. “The importance of industry links in merger waves.” *Journal of Finance*, 69(2): 527–576.
- Akerlof, George A, and Rachel E Kranton.** 2000. “Economics and identity.” *Quarterly Journal of Economics*, 115(3): 715–753.
- Alesina, Alberto F, Francesca Lotti, and Paolo Emilio Mistrulli.** 2013. “Do women pay more for credit? Evidence from Italy.” *Journal of the European Economic Association*, 11(1): 45–66.
- Almeida, Heitor, Chang-Soo Kim, and Hwanki Brian Kim.** 2015. “Internal capital markets in business groups: Evidence from the Asian financial crisis.” *Journal of Finance*, 70(6): 2539–2586.
- Amihud, Yakov.** 2002. “Illiquidity and stock returns: cross-section and time-series effects.” *Journal of Financial Markets*, 5(1): 31–56.
- Anjos, Fernando, and Cesare Fracassi.** 2015. “Shopping for information? Diversification and the network of industries.” *Management Science*, 61(1): 161–183.

- Appel, Ian R, Todd A Gormley, and Donald B Keim.** 2016. "Passive investors, not passive owners." *Journal of Financial Economics*, 121(1): 111–141.
- Arnold, David, Will Dobbie, and Crystal S Yang.** 2018. "Racial bias in bail decisions." *Quarterly Journal of Economics*.
- Arrow, Kenneth.** 1973. "The theory of discrimination." *Discrimination in Labor Markets*, 3(10): 3–33.
- Autor, David, and David Dorn.** 2009. "This job is "getting old": measuring changes in job opportunities using occupational age structure." *American Economic Review*, 99(2): 45–51.
- Azar, José, Martin C Schmalz, and Isabel Tecu.** 2018. "Anticompetitive effects of common ownership." *Journal of Finance*, 73(4): 1513–1565.
- Bae, Gil S, Youngsoon S Cheon, and Jun-Koo Kang.** 2008. "Intragroup propping: Evidence from the stock-price effects of earnings announcements by Korean business groups." *Review of Financial Studies*, 21(5): 2015–2060.
- Baghai, Ramin, Rui Silva, Viktor Thell, and Vikrant Vig.** 2018. "Talent in distressed firms: Investigating the labor costs of financial distress." *Working Paper*.
- Bagnoli, Mark, William Kross, and Susan G Watts.** 2002. "The information in management's expected earnings report date: A day late, a penny short." *Journal of Accounting Research*, 40(5): 1275–1296.
- Bai, John Jianqiu, Douglas J Fairhurst, and Matthew Serfling.** 2018. "Employment protection, investment, and firm growth." *Working Paper*.
- Bailey, Warren, G Andrew Karolyi, and Carolina Salva.** 2006. "The economic consequences of increased disclosure: Evidence from international cross-listings." *Journal of Financial Economics*, 81(1): 175–213.
- Barber, Brad M, and Terrance Odean.** 2001. "Boys will be boys: Gender, overconfidence, and common stock investment." *Quarterly Journal of Economics*, 116(1): 261–292.
- Barber, Brad M, Anna Scherbina, and Bernd Schlusche.** 2017. "Performance isn't everything: Personal characteristics and career outcomes of mutual fund managers." *Working Paper*.
- Barinov, Alexander, Shawn Saeyeul Park, and Celim Yildizhan.** 2018. "Firm complexity and post-earnings-announcement drift." *Working Paper*.
- Bartov, Eli, Suresh Radhakrishnan, and Itzhak Krinsky.** 2000. "Investor sophistication and patterns in stock returns after earnings announcements." *The Accounting Review*, 75(1): 43–63.
- Battalio, Robert H, and Richard R Mendenhall.** 2005. "Earnings expectations, investor trade size, and anomalous returns around earnings announcements." *Journal of Financial Economics*, 77(2): 289–319.
- Becker, Gary S.** 1957. "The economics of discrimination." *University of Chicago*.

- Becker, Gary S.** 1993. “Nobel lecture: The economic way of looking at behavior.” *Journal of Political Economy*, 101(3): 385–409.
- Begley, Joy, and Paul E Fischer.** 1998. “Is there information in an earnings announcement delay?” *Review of Accounting Studies*, 3(4): 347–363.
- Bellucci, Andrea, Alexander Borisov, and Alberto Zazzaro.** 2010. “Does gender matter in bank-firm relationships? Evidence from small business lending.” *Journal of Banking & Finance*, 34(12): 2968–2984.
- Bena, Jan, and Kai Li.** 2014. “Corporate innovations and mergers and acquisitions.” *Journal of Finance*, 69(5): 1923–1960.
- Bernard, Andrew B, Emily J Blanchard, Ilke Van Beveren, and Hylke Vandenbussche.** 2018. “Carry-along trade.” *The Review of Economic Studies*, 86(2): 526–563.
- Bernard, Andrew B, J Bradford Jensen, Stephen J Redding, and Peter K Schott.** 2007. “Firms in international trade.” *Journal of Economic perspectives*, 21(3): 105–130.
- Bernard, Andrew B, Stephen J Redding, and Peter K Schott.** 2010. “Multiple-product firms and product switching.” *American Economic Review*, 100(1): 70–97.
- Bernard, Victor L, and Jacob K Thomas.** 1989. “Post-earnings-announcement drift: delayed price response or risk premium?” *Journal of Accounting Research*, 27: 1–36.
- Bernard, Victor L, and Jacob K Thomas.** 1990. “Evidence that stock prices do not fully reflect the implications of current earnings for future earnings.” *Journal of Accounting and Economics*, 13(4): 305–340.
- Bertrand, Marianne.** 2011. “New perspectives on gender.” *Handbook of Labor Economics*, 4: 1543–1590.
- Bertrand, Marianne, and Sendhil Mullainathan.** 2004. “Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination.” *American Economic Review*, 94(4): 991–1013.
- Bertrand, Marianne, Sandra E Black, Sissel Jensen, and Adriana Lleras-Muney.** 2014. “Breaking the glass ceiling? The effect of board quotas on female labor market outcomes in Norway.” *Review of Economic Studies*.
- Bhattacharya, Nilabhra.** 2001. “Investors’ trade size and trading responses around earnings announcements: An empirical investigation.” *The Accounting Review*, 76(2): 221–244.
- Bhushan, Ravi.** 1994. “An informational efficiency perspective on the post-earnings announcement drift.” *Journal of Accounting and Economics*, 18(1): 45–65.
- Blatter, Marc, Samuel Muehleman, and Samuel Schenker.** 2012. “The costs of hiring skilled workers.” *European Economic Review*, 56(1): 20–35.
- Bloom, Nicholas, Mark Schankerman, and John Van Reenen.** 2013. “Identifying technology spillovers and product market rivalry.” *Econometrica*, 81(4): 1347–1393.
- Boehm, Johannes, Swati Dhingra, and John Morrow.** 2019. “The comparative advantage of firms.” *Working Paper*.

- Bohren, J Aislinn, Alex Imas, and Michael Rosenberg.** 2017. "The dynamics of discrimination: Theory and evidence." *Working Paper*.
- Bolton, Patrick, and Ernst-Ludwig Von Thadden.** 1998. "Blocks, liquidity, and corporate control." *Journal of Finance*, 53(1): 1–25.
- Bordalo, Pedro, Katherine B Coffman, Nicola Gennaioli, and Andrei Shleifer.** 2017. "Beliefs about gender." *Working Paper*.
- Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer.** 2016. "Stereotypes." *Quarterly Journal of Economics*, 131(4): 1753–1794.
- Bottazzi, Laura, Marco Da Rin, and Thomas Hellmann.** 2016. "The importance of trust for investment: Evidence from venture capital." *Review of Financial Studies*, 29(9): 2283–2318.
- Boulland, Romain, and Olivier Dessaint.** 2017. "Announcing the announcement." *Journal of Banking & Finance*, 82: 59–79.
- Bradshaw, Mark T, Scott A Richardson, and Richard G Sloan.** 2001. "Do analysts and auditors use information in accruals?" *Journal of Accounting research*, 39(1): 45–74.
- Breda, Thomas, and Son Thierry Ly.** 2015. "Professors in core science fields are not always biased against women: Evidence from France." *American Economic Journal: Applied Economics*, 7(4): 53–75.
- Cao, Jie, Tarun Chordia, and Chen Lin.** 2016. "Alliances and return predictability." *Journal of Financial and Quantitative Analysis*, 51(5): 1689–1717.
- Casamatta, Catherine.** 2003. "Financing and advising: optimal financial contracts with venture capitalists." *Journal of Finance*, 58(5): 2059–2085.
- Chandler, Alfred D.** 1992. "Organizational capabilities and the economic history of the industrial enterprise." *Journal of Economic Perspectives*, 6(3): 79–100.
- Chari, Anusha, and Paul Goldsmith-Pinkham.** 2017. "Gender representation in economics across topics and time: Evidence from the NBER summer institute." *Working Paper*.
- Chen, Deqiu, Huasheng Gao, and Yujing Ma.** 2018. "Human capital driven acquisition: Evidence from the inevitable disclosure doctrine." *Working Paper*.
- Cheng, Wenya, and John Morrow.** 2018. "Firm productivity differences from factor markets." *Journal of Industrial Economics*, 66(1): 126–171.
- Claessens, Stijn, Simeon Djankov, and Larry HP Lang.** 2000. "The separation of ownership and control in East Asian corporations." *Journal of Financial Economics*, 58(1-2): 81–112.
- Coffman, Katherine B, Christine L Exley, and Muriel Niederle.** 2018. "When gender discrimination is not about gender." *Working Paper*.
- Cohen, Lauren, and Andrea Frazzini.** 2008. "Economic links and predictable returns." *Journal of Finance*, 63(4): 1977–2011.
- Cohen, Lauren, and Dong Lou.** 2012. "Complicated firms." *Journal of Financial Economics*, 104(2): 383–400.

- Cook, Cody, Rebecca Diamond, Jonathan Hall, John A List, Paul Oyer, et al.** 2018. “The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers.” *Working Paper*.
- Cornell, Bradford, and Ivo Welch.** 1996. “Culture, information, and screening discrimination.” *Journal of Political Economy*, 104(3): 542–571.
- Da Rin, Marco, Thomas Hellmann, and Manju Puri.** 2013. “A survey of venture capital research.” *Handbook of the Economics of Finance*, 2: 573–648.
- Davis, Steven J, and John Haltiwanger.** 1992. “Gross job creation, gross job destruction, and employment reallocation.” *Quarterly Journal of Economics*, 107(3): 819–863.
- DeFond, Mark, Mingyi Hung, and Robert Trezevant.** 2007. “Investor protection and the information content of annual earnings announcements: International evidence.” *Journal of Accounting and Economics*, 43(1): 37–67.
- Del Carpio, Lucia, and Maria Guadalupe.** 2018. “More women in tech? Evidence from a field experiment addressing social identity.” *Working Paper*.
- DellaVigna, Stefano, and Joshua M Pollet.** 2009. “Investor inattention and Friday earnings announcements.” *Journal of Finance*, 64(2): 709–749.
- Dessaint, Olivier, Andrey Golubov, and Paolo Volpin.** 2017. “Employment protection and takeovers.” *Journal of Financial Economics*, 125(2): 369–388.
- Djankov, Simeon, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer.** 2008. “The law and economics of self-dealing.” *Journal of Financial Economics*, 88(3): 430–465.
- Dobbie, Will, Andres Liberman, Daniel Paravisini, and Vikram Pathania.** 2018. “Measuring bias in consumer lending.” *Working Paper*.
- Doyle, Jeffrey T, and Matthew J Magilke.** 2009. “The timing of earnings announcements: An examination of the strategic disclosure hypothesis.” *The Accounting Review*, 84(1): 157–182.
- Dyck, Alexander, and Luigi Zingales.** 2004. “Private benefits of control: An international comparison.” *Journal of Finance*, 59(2): 537–600.
- Eagly, Alice H, and Steven J Karau.** 2002. “Role congruity theory of prejudice toward female leaders.” *Psychological Review*, 109(3): 573.
- Eckbo, Espen.** 1983. “Horizontal mergers, collusion, and stockholder wealth.” *Journal of financial Economics*, 11(1-4): 241–273.
- Egan, Mark L, Gregor Matvos, and Amit Seru.** 2017. “When Harry fired Sally: The double standard in punishing misconduct.” *Working Paper*.
- Engelberg, Joseph, and Pengjie Gao.** 2011. “In search of attention.” *Journal of Finance*, 66(5): 1461–1499.
- Ewens, Michael, and Richard R Townsend.** 2017. “Are early stage investors biased against women?” *Working Paper*.
- Faccio, Mara, and Larry HP Lang.** 2002. “The ultimate ownership of Western European corporations.” *Journal of Financial Economics*, 65(3): 365–395.

- Faccio, Mara, and William O'Brien.** 2017. "Business groups and employment." *Working Paper*.
- Fan, Joseph PH, and Tak Jun Wong.** 2002. "Corporate ownership structure and the informativeness of accounting earnings in East Asia." *Journal of Accounting and Economics*, 33(3): 401–425.
- Fan, Joseph P. H., and Vidhan Goyal.** 2006. "On the patterns and wealth effects of vertical mergers." *Journal of Business*, 79(2): 877–902.
- Fan, Joseph PH, Li Jin, and Guojian Zheng.** 2016. "Revisiting the bright and dark sides of capital flows in business groups." *Journal of Business Ethics*, 134(4): 509–528.
- Farrell, Joseph, and Carl Shapiro.** 1990. "Horizontal mergers: an equilibrium analysis." *American Economic Review*, 107–126.
- Fraisse, Henri, Francis Kramarz, and Corinne Prost.** 2015. "Labor disputes and job flows." *ILR Review*, 68(5): 1043–1077.
- Frederickson, James R, and Leon Zolotoy.** 2015. "Competing earnings announcements: Which announcement do investors process first?" *The Accounting Review*, 91(2): 441–462.
- Gennaioli, Nicola, and Andrei Shleifer.** 2010. "What comes to mind." *Quarterly Journal of Economics*, 125(4): 1399–1433.
- Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer.** 2013. "Human capital and regional development." *Quarterly Journal of Economics*, 128(1): 105–164.
- Gibbons, Robert, and Michael Waldman.** 2004. "Task-specific human capital." *American Economic Review*, 94(2): 203–207.
- Gilje, Erik P, Todd A Gormley, and Doron Levit.** 2018. "The rise of common ownership." *Working Paper*.
- Goldin, Claudia, and Cecilia Rouse.** 2000. "Orchestrating impartiality: The impact of "blind" auditions on female musicians." *American Economic Review*, 90(4): 715–741.
- Golubov, Andrey, Alfred Yawson, and Huizhong Zhang.** 2015. "Extraordinary acquirers." *Journal of Financial Economics*, 116(2): 314–330.
- Gompers, Paul A, and Sophie Q Wang.** 2017a. "And the children shall lead: Gender diversity and performance in venture Capital." *Working Paper*.
- Gompers, Paul A, and Sophie Q Wang.** 2017b. "Diversity in innovation." *Working Paper*.
- Gompers, Paul, and Josh Lerner.** 1999. "An analysis of compensation in the US venture capital partnership." *Journal of Financial Economics*, 51(1): 3–44.
- Gompers, Paul A, Vladimir Mukharlyamov, Emily Weisburst, and Yuhai Xuan.** 2014. "Gender effects in venture capital." *Working Paper*.
- Gornall, Will, and Ilya A Strebulaev.** 2019. "Gender, race, and entrepreneurship: A randomized field experiment on venture capitalists and angels." *Working paper*.

- Grossman, Sanford J, and Oliver D Hart.** 1986. “The costs and benefits of ownership: a theory of vertical and lateral integration.” *Journal of Political economy*, 94(4): 691–719.
- Haltiwanger, John, Ron S Jarmin, and Javier Miranda.** 2013. “Who creates jobs? Small versus large versus young.” *Review of Economics and Statistics*, 95(2): 347–361.
- Hart, Oliver, and John Moore.** 1990. “Property rights and the nature of the firm.” *Journal of Political Economy*, 98(6): 1119–1158.
- Hellmann, Thomas, and Manju Puri.** 2000. “The interaction between product market and financing strategy: The role of venture capital.” *Review of Financial Studies*, 13(4): 959–984.
- Hellmann, Thomas, and Manju Puri.** 2002. “Venture capital and the professionalization of start-up firms: Empirical evidence.” *Journal of Finance*, 57(1): 169–197.
- Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh.** 2009. “Driven to distraction: Extraneous events and underreaction to earnings news.” *Journal of Finance*, 64(5): 2289–2325.
- Hoberg, Gerard, and Gordon Phillips.** 2010. “Product market synergies and competition in mergers and acquisitions: A text-based analysis.” *Review of Financial Studies*, 23(10): 3773–3811.
- Hoberg, Gerard, and Gordon Phillips.** 2016. “Text-based network industries and endogenous product differentiation.” *Journal of Political Economy*, 124(5): 1423–1465.
- Holderness, Clifford G.** 2007. “The myth of diffuse ownership in the United States.” *Review of Financial Studies*, 22(4): 1377–1408.
- Hombert, Johan, Antoinette Schoar, David Sraer, and David Thesmar.** 2017. “Can unemployment insurance spur entrepreneurial activity? Evidence from France.” *Working Paper*.
- Howell, Sabrina, and Ramana Nanda.** 2019. “Networking frictions in venture capital, and the gender gap in entrepreneurship.” *Working Paper*.
- Howell, Sabrina T.** 2017. “Financing innovation: Evidence from R&D grants.” *American Economic Review*, 107(4): 1136–1164.
- Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow.** 2013. “The allocation of talent and US economic growth.” *Working Paper*.
- Huang, Xing.** 2015. “Thinking outside the borders: Investors’ underreaction to foreign operations information.” *Review of Financial Studies*, 28(11): 3109–3152.
- Hung, Mingyi, Xi Li, and Shiheng Wang.** 2014. “Post-earnings-announcement drift in global markets: Evidence from an information shock.” *Review of Financial Studies*, 28(4): 1242–1283.
- Hurst, Erik, and Benjamin Wild Pugsley.** 2011. “What do small businesses do?” *Brookings Papers on Economic Activity*, 73.
- Hvide, Hans K, and Georgios A Panos.** 2014. “Risk tolerance and entrepreneurship.” *Journal of Financial Economics*, 111(1): 200–223.

- Jannati, Sima, Alok Kumar, Alexandra Niessen-Ruenzi, and Justin Wolfers.** 2016. "In-group bias in financial markets." *Working Paper*.
- Johnson, Simon, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer.** 2000. "Tunneling." *American Economic Review*, 90(2): 22–27.
- Johnson, Travis L, and Eric C So.** 2018. "Time will tell: Information in the timing of scheduled earnings news." *Journal of Financial and Quantitative Analysis*, 53(6): 2431–2464.
- Jouini, Elyes, Paul Karehnke, and Clotilde Napp.** 2018. "Self-confidence, self-esteem, and achievement-related choices." *Journal of Economic Behavior & Organization*, 148: 34–45.
- Kaniel, Ron, Shuming Liu, Gideon Saar, and Sheridan Titman.** 2012. "Individual investor trading and return patterns around earnings announcements." *Journal of Finance*, 67(2): 639–680.
- Kaplan, Steven N, and Antoinette Schoar.** 2005. "Private equity performance: Returns, persistence, and capital flows." *Journal of Finance*, 60(4): 1791–1823.
- Kaplan, Steven N, and Per Stromberg.** 2001. "Venture capitals as principals: Contracting, screening, and monitoring." *American Economic Review*, 91(2): 426–430.
- Kaplan, Steven N, and Per Strömberg.** 2003. "Financial contracting theory meets the real world: An empirical analysis of venture capital contracts." *Review of Economic Studies*, 70(2): 281–315.
- Kauffman, Foundation.** 2017. "Kauffman index of start-up activity: National trends." *Report*.
- Ke, Bin, and Santhosh Ramalingegowda.** 2005. "Do institutional investors exploit the post-earnings announcement drift?" *Journal of Accounting and Economics*, 39(1): 25–53.
- Kerr, Sari Pekkala, William R Kerr, and Tina Xu.** 2017. "Personality traits of entrepreneurs: A review of recent literature." *Working Paper*.
- Kerr, William R, Josh Lerner, and Antoinette Schoar.** 2011. "The consequences of entrepreneurial finance: Evidence from angel financings." *Review of Financial Studies*, 27(1): 20–55.
- Kumar, Alok.** 2010. "Self-selection and the forecasting abilities of female equity analysts." *Journal of Accounting Research*, 48(2): 393–435.
- Kumar, Alok, Alexandra Niessen-Ruenzi, and Oliver G Spalt.** 2015. "What's in a name? Mutual fund flows when managers have foreign-sounding names." *Review of Financial Studies*, 28(8): 2281–2321.
- Laeven, Luc, and Ross Levine.** 2007. "Complex ownership structures and corporate valuations." *Review of Financial Studies*, 21(2): 579–604.
- Lagaras, Spyridon.** 2017. "Corporate takeovers and labor restructuring." *Working Paper*.
- Lamont, Owen A, and Richard H Thaler.** 2003. "Can the market add and subtract? Mispricing in tech stock carve-outs." *Journal of Political Economy*, 111(2): 227–268.

- Landier, Augustin, and David Thesmar.** 2008. “Financial contracting with optimistic entrepreneurs.” *Review of Financial Studies*, 22(1): 117–150.
- Lee, Kyeong Hun, David C. Mauer, and Emma Qianying Xu.** 2018. “Human capital relatedness and mergers and acquisitions.” *Journal of Financial Economics*, 129(1): 111–135.
- Levine, Ross, and Yona Rubinstein.** 2017. “Smart and illicit: Who becomes an entrepreneur and do they earn more?” *Quarterly Journal of Economics*, 132(2): 963–1018.
- Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan.** 2018. “Measuring corporate culture using machine learning.” Working Paper.
- Madsen, Joshua.** 2017. “Anticipated earnings announcements and the customer–supplier anomaly.” *Journal of Accounting Research*, 55(3): 709–741.
- Maksimovic, Vojislav, and Gordon M Phillips.** 2013. “Conglomerate firms, internal capital markets, and the theory of the firm.” *Annual Review of Financial Economics*, 5(1): 225–244.
- Maksimovic, Vojislav, and Gordon Phillips.** 2002. “Do conglomerate firms allocate resources inefficiently across industries? Theory and evidence.” *Journal of Finance*, 57(2): 721–767.
- Maksimovic, Vojislav, Gordon Phillips, and Nagpurnanand R Prabhala.** 2011. “Post-merger restructuring and the boundaries of the firm.” *Journal of Financial Economics*, 102(2): 317–343.
- Massa, Massimo, and Alminas Žaldokas.** 2017. “Information transfers among co-owned firms.” *Journal of Financial Intermediation*, 31: 77–92.
- Massa, Massimo, and Zahid Rehman.** 2008. “Information flows within financial conglomerates: Evidence from the banks–mutual funds relation.” *Journal of Financial Economics*, 89(2): 288–306.
- Matsa, David A.** 2010. “Capital structure as a strategic variable: evidence from collective bargaining.” *Journal of Finance*, 65(3): 1197–1232.
- Matsa, David A, and Amalia R Miller.** 2013. “A female style in corporate leadership? Evidence from quotas.” *American Economic Journal: Applied Economics*, 5(3): 136–69.
- Matsusaka, John G.** 2001. “Corporate diversification, value maximization, and organizational capabilities.” *Journal of Business*, 74(3): 409–431.
- Maug, Ernst.** 1998. “Large shareholders as monitors: Is there a trade-off between liquidity and control?” *Journal of Finance*, 53(1): 65–98.
- Ma, Wenting, Paige Ouimet, and Elena Simintzi.** 2016. “Mergers and acquisitions, technological change and inequality.” Working Paper.
- McCardle, Kevin F, and S Viswanathan.** 1994. “The direct entry versus takeover decision and stock price performance around takeovers.” *Journal of Business*, 1–43.
- McElroy, Marjorie B.** 2016. “Report: Committee on the status of women in the economics profession (CSWEP).” *American Economic Review*, 106(5): 750–73.

- Meier, Kristina, Alexandra Niessen-Ruenzi, and Stefan Ruenzi.** 2017. "The impact of role models on women's self-selection in competitive environments." *Working Paper*.
- Merton, Robert C.** 1987. "A simple model of capital market equilibrium with incomplete information." *Journal of Finance*, 42(3): 483–510.
- Merz, Monika, and Eran Yashiv.** 2007. "Labor and the market value of the firm." *American Economic Review*, 97(4): 1419–1431.
- Michaely, Roni, Amir Rubin, and Alexander Vedrashko.** 2013. "Corporate governance and the timing of earnings announcements." *Review of Finance*, 18(6): 2003–2044.
- MIWE.** 2018. "Mastercard index of women entrepreneurs." *Report*.
- Moretti, Enrico.** 2010. "Local labor markets." *National Bureau of Economic Research*.
- Moskowitz, Tobias J, and Annette Vissing-Jorgensen.** 2002. "The returns to entrepreneurial investment: A private equity premium puzzle?" *American Economic Review*, 92(4): 745–778.
- Ng, Jeffrey, Tjomme O Rusticus, and Rodrigo S Verdi.** 2008. "Implications of transaction costs for the post-earnings announcement drift." *Journal of Accounting Research*, 46(3): 661–696.
- Niederle, Muriel, and Lise Vesterlund.** 2007. "Do women shy away from competition? Do men compete too much?" *Quarterly Journal of Economics*, 122(3): 1067–1101.
- Niessen-Ruenzi, Alexandra, and Stefan Ruenzi.** 2018. "Sex matters: Gender bias in the mutual fund industry." *Management Science*.
- Nocke, Volker, and Stephen Yeaple.** 2007. "Cross-border mergers and acquisitions vs. greenfield foreign direct investment: The role of firm heterogeneity." *Journal of International Economics*, 72(2): 336–365.
- Ors, Evren, Frédéric Palomino, and Eloic Peyrache.** 2013. "Performance gender gap: Does competition matter?" *Journal of Labor Economics*, 31(3): 443–499.
- Ouimet, Paige Parker, and Rebecca Zarutskie.** 2016. "Acquiring labor." Working Paper.
- Penrose, Edith.** 1955. "Limits to the growth and size of firms." *American Economic Review*, 45(2): 531–543.
- Phelps, Edmund S, et al.** 1972. "The statistical theory of racism and sexism." *American Economic Review*, 62(4): 659–661.
- Phillips, Gordon M, and Alexei Zhdanov.** 2013. "R&D and the Incentives from Merger and Acquisition Activity." *The Review of Financial Studies*, 26(1): 34–78.
- Prado, Melissa, Pedro AC Saffi, and Jason Sturgess.** 2016. "Ownership structure, limits to arbitrage, and stock returns: Evidence from equity lending markets." *Review of Financial Studies*, 29(12): 3211–3244.
- Puri, Manju, and David T Robinson.** 2013. "The economic psychology of entrepreneurship and family business." *Journal of Economics & Management Strategy*, 22(2): 423–444.

- Puri, Manju, and Rebecca Zarutskie.** 2012. "On the life cycle dynamics of venture-capital-and non-venture-capital-financed firms." *Journal of Finance*, 67(6): 2247–2293.
- Raina, Sahil.** 2017. "VC financing and the entrepreneurship gender gap." *Working Paper*.
- Rajan, Raghuram G, and Luigi Zingales.** 1998. "Financial dependence and growth." *American Economic Review*, 559–586.
- Rajan, Raghuram, Henri Servaes, and Luigi Zingales.** 2000. "The cost of diversity: The diversification discount and inefficient investment." *Journal of Finance*, 55(1): 35–80.
- Ramnath, Sundaresh.** 2002. "Investor and analyst reactions to earnings announcements of related firms: An empirical analysis." *Journal of Accounting Research*, 40(5): 1351–1376.
- Reuben, Ernesto, Paola Sapienza, and Luigi Zingales.** 2014. "How stereotypes impair women's careers in science." *Proceedings of the National Academy of Sciences*, 111(12): 4403–4408.
- Rhodes-Kropf, Matthew, and David T. Robinson.** 2008. "The market for mergers and the boundaries of the firm." *Journal of Finance*, 63(3): 1169–1211.
- Robb, Alicia M, and David T Robinson.** 2014. "The capital structure decisions of new firms." *Review of Financial Studies*, 27(1): 153–179.
- Romanski, Hilton.** 2017. "Cisco's 200th Acquisition - A Tradition of Advancement, Disruption and Growth." *Cisco Blog*.
- Sapienza, Paola, Luigi Zingales, and Dario Maestripieri.** 2009. "Gender differences in financial risk aversion and career choices are affected by testosterone." *Proceedings of the National Academy of Sciences*, 106(36): 15268–15273.
- Sarsons, Heather.** 2017. "Gender differences in recognition for group work." *Working Paper*.
- Schoar, Antoinette.** 2010. "The divide between subsistence and transformational entrepreneurship." *Innovation Policy and the Economy*, 10(1): 57–81.
- Serfling, Matthew.** 2016. "Firing costs and capital structure decisions." *Journal of Finance*, 71(5): 2239–2286.
- Simintzi, Elena, Vikrant Vig, and Paolo Volpin.** 2014. "Labor protection and leverage." *Review of Financial Studies*, 28(2): 561–591.
- Slovin, Myron B, and Marie E Sushka.** 1997. "The implications of equity issuance decisions within a parent-subsidiary governance structure." *Journal of Finance*, 52(2): 841–857.
- Sørensen, Morten.** 2007. "How smart is smart money? A two-sided matching model of venture capital." *Journal of Finance*, 62(6): 2725–2762.
- Stein, Jeremy C.** 1989. "Efficient capital markets, inefficient firms: A model of myopic corporate behavior." *Quarterly Journal of Economics*, 104(4): 655–669.
- Stein, Jeremy C.** 1997. "Internal capital markets and the competition for corporate resources." *Journal of Finance*, 52(1): 111–133.

- Tate, Geoffrey A, and Liu Yang.** 2016a. "The human factor in acquisitions: cross-industry labor mobility and corporate diversification." *Working Paper*.
- Tate, Geoffrey, and Liu Yang.** 2016b. "The bright side of corporate diversification: evidence from internal labor markets." *Review of Financial Studies*, 28(8): 2203–2249.
- Tversky, Amos, and Daniel Kahneman.** 1983. "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment." *Psychological Review*, 90(4): 293.
- Walther, Beverly R.** 1997. "Investor sophistication and market earnings expectations." *Journal of Accounting Research*, 35(2): 157–179.
- Winton, Andrew, and Vijay Yerramilli.** 2008. "Entrepreneurial finance: Banks versus venture capital." *Journal of Financial Economics*, 88(1): 51–79.
- Wysocki, Bernard Jr.** 1997. "Many mergers driven by search for fresh talent." *Wall Street Journal*.
- Xu, Sheng-Jun.** 2018. "Skilled labor supply and corporate investment: Evidence from the H-1B visa program." *Working Paper*.
- Yip, George S.** 1982. "Diversification entry: Internal development versus acquisition." *Strategic Management Journal*, 3(4): 331–345.

## RÉSUMÉ

---

Cette thèse comprend trois chapitres et étudie la structure organisationnelle de l'entreprise à des stades différents de son cycle de vie: démarrage, croissance, grande entreprise. Le premier chapitre examine les raisons sous-jacentes aux différences de financement entre hommes et femmes entrepreneurs dans le contexte de l'industrie du capital risque. Ce chapitre met en lumière les effets des stéréotypes de genre qui entravent la croissance des jeunes entreprises fondées par des entrepreneurs issus des minorités de genre. Dans ce contexte, le capital humain des entrepreneurs atténue dans une certaine mesure les stéréotypes des investisseurs. Le deuxième chapitre décrit les conditions dans lesquelles les entreprises choisissent de croître en achetant une entreprise existante plutôt que de se développer à partir des ressources préexistantes en capital humain de l'entreprise. Le troisième chapitre porte sur les grands groupes d'entreprises. Ce chapitre montre que les investisseurs ne sont pas toujours conscients des limites de la structure de l'entreprise et omettent de l'information prédictive révélée à un autre niveau du groupe.

## MOTS CLÉS

---

Finance d'Entreprise, Entrepreneuriat, Croissance, Capital Risque, Capital Humain, Genre

## ABSTRACT

---

This thesis consists of three chapters and studies the firm's organizational structure at a different stage of its life cycle: early-stage, growth, business group. The first chapter investigates the underlying reasons for the gender funding gap in the venture capital industry. It highlights a significant role for investors' stereotypes that ultimately impedes minority-founded startups' growth. Entrepreneurs' human capital mitigates to some extent investors' stereotypes. The second chapter identifies conditions under which firms choose to grow by buying an incumbent company as opposed to building on their pre-existing human capital resources. The third chapter focuses on large business groups. It provides evidence that investors are not always aware of the boundaries of the firm and miss predictive information released at another level of the group.

## KEYWORDS

---

Corporate Finance, Entrepreneurship, Growth, Venture Capital, Human Capital, Gender