Essays on Strategy, Geography, and Firm Performance

A dissertation presented

by

Jasmina Beganović Chauvin

to

The Strategy Unit at the Harvard Business School

in partial fulfillment of the requirements

for the degree of

Doctor of Business Administration

in the subject of

Strategy

Harvard University

Cambridge, Massachusetts

April 2018

Dissertation Advisor:
Professor Juan Alcácer

Author:
Jasmina Beganović Chauvin

Essays on Strategy, Geography, and Firm Performance

Abstract

Corporate strategy — what activities a firm performs — and location strategy — where it performs them — have mostly been studied separately. However, geographic proximity enables the exchange of goods, workers, and knowledge — precisely the types of resource flows that also motivate internalizing activities inside firm boundaries. This dissertation presents three essays that explore the interdependence of corporate- and location strategy and its implications for location choice and firm performance.

Chapter 1 studies the effects of geographic proximity between firms in the same industry on their survival. While in theory, co-location can enhance firm productivity, the existing empirical evidence is mixed. In this paper, I argue that proximity between firms affects their performance differently depending on whether they compete locally or in broader national markets. Using road upgrades in the context of Brazil as an exogenous shock to proximity between incumbent firms, I find that in locally traded industries, greater proximity leads to increased exit of the smallest firms and higher survival rates of the largest — effects that are consistent with increased competition. Meanwhile in nationally traded industries, firms of all sizes see improved survival rates, consistent with increased agglomeration spillovers. The results shed light on contradictory findings in the literature and show how investments in transportation infrastructure, such as roads, intensify both competition and agglomeration forces.

In contrast to the focus on stand-alone firms in Chapter 1, Chapter 2 studies the spatial organization of complex, multibusiness firms. While prior research has focused on how firms co-locate with others, here we focus on the geographic proximity between the different

units of the firm itself. We propose and test the hypothesis that multibusiness firms exhibit "internal agglomeration" — a systematic co-location of their different plants — and that this is driven by the desire to share resources within the firm. Using data on the location and corporate structure of a large sample of U.S. manufacturing firms, we find that internal agglomeration exists and is primarily related to the sharing of labor. The findings suggest that internal labor markets are potentially an important source of competitive advantage in multibusiness firms.

Building on the findings of Chapter 2, Chapter 3 studies the extent and drivers of internal labor markets in multibusiness firms directly. Using a large sample of multi-business firms from Brazil and a rich employer-employee matched dataset, we track all internal worker movements across the firms' units. We find that multibusiness firms redeploy a large share of their workers internally, especially managers and workers with more firm-specific experience. Redeployed workers earn a large wage premium over otherwise comparable workers hired though external labor markets. Geographic proximity and resource relatedness between the firms' plants facilitate redeployment. In contrast to prevailing views of internal labor markets as a means to avoid external labor market frictions, our findings are consistent with internal labor markets as conduits of knowledge.

Taken together, the three chapters of the dissertation provide evidence that strategic decisions around a firm's product- and geographic boundaries are intimately related, and that resource sharing is implicated in both.

Contents

Abstract iii

Introduction **1**

1 When Distance Shrinks: The Effects of Competitor Proximity on Firm Survival **5**

1.1 Introduction 5
1.2 Conceptual Framework 9
1.3 Data and Measurement 16
1.4 Identification Strategy and Institutional Context 25
1.5 Empirical Specification 30
1.6 Results 33
1.7 Conclusion 45

2 The Internal Agglomeration of Multibusiness Firms **47**

2.1 Introduction 47
2.2 Theoretical Background and Hypotheses 53
2.3 Data and Methodology 60
2.4 Results 66
2.5 Further Tests and Robustness Checks 72
2.6 Discussion and Conclusion 74

3 Internal Labor Markets in Multibusiness Firms **78**

3.1 Introduction 78
3.2 Theory and Hypotheses 83
3.3 Data and Empirical Strategy 91
3.4 Results 96
3.5 Conclusion 103

References **105**

Appendix A **113**
A.1 Supplementary Tables for Chapter 1 113
A.2 Supplementary Figures for Chapter 1 117
A.3 Supplementary Tables for Chapter 2 124
A.4 Supplementary Figures for Chapter 2 125
A.5 Supplementary Figures for Chapter 3 128

List of Tables

1.1 Locally Traded Industries: Effects of Change in Co-Location on Firm Survival during 2007–2014 . . . 35
1.2 Nationally Traded Industries: Effects of Change in Co-Location on Firm Survival, 2007–2014 . . . 39
1.3 Effects of Change in Co-Location on Relocations and Product Switches . . . 41
1.4 Robustness Checks . . . 44

2.1 Firms and Establishments Included in the Analysis . . . 61
2.2 Summary Statistics . . . 67
2.3 Results of T-Tests of Equality of Means . . . 68
2.4 OLS results of Univariate Regressions of Internal and External Agglomeration on Industry Relatedness . . . 69
2.5 OLS results of Multivariate Regressions of Internal and External Agglomeration on Industry Relatedness . . . 70
2.6 OLS results of the Difference Between Internal and External Agglomeration 72
2.7 Robustness Checks . . . 76
2.8 Correlations of Main Variables . . . 77

3.1 Worker Summary Statistics . . . 94
3.2 Establishment Summary Statistics . . . 94
3.3 Destination-Origin Dyad Summary Statistics . . . 94
3.4 Percentage of Workers Redeployed by Establishment . . . 96
3.5 Percentage of Workers Redeployed by Occupation . . . 97
3.6 Redeployment of Workers . . . 98
3.7 Wages of Redeployed Workers . . . 100
3.8 Adjustment Costs and Redeployment . . . 101
3.9 Adjustment Costs and Redeployment When Closing . . . 102

A.1 Firm Descriptive Statistics in 2007 . . . 113
A.2 Federal Roads by Surface Type in 2007 and 2014 . . . 114
A.3 Federal Roads by Surface Condition in 2007 and 2014 . . . 114

A.4 Assumed Velocities (km/hr) Given Road Surface and Condition 114
A.5 Fifteen Largest Industries (by Number of Firms in Sample) 115
A.6 Correlations Between the Key Variables 116
A.7 Most Agglomerated Industry Pairs, Multibusiness (top) and Stand-alone (bottom) 124

List of Figures

1.1 Predicted Effects of Increases in Co-Location in Locally Traded and Nationally Traded Industries . . . 16
1.2 Change in Local Travel Times Between in Brazilian Municipalities, 2007–2014 23
1.3 Pre-period Survival Rates in Locally- and Nationally Traded Industries . . . 30
1.4 Origin and Destination Co-Location for Firms that Move and Switch 42

A.1 Examples of Markets as Defined by Four Hours of Travel Time 117
A.2 Programa de Aceleração do Crescimento — Programmed Investments 118
A.3 Distances and Travel Times in Google Maps and O-D matrix 119
A.4 Distribution of Distances Globally and Locally 119
A.5 Median Distance Manufacturing Goods Shipped by Truck, by NAICS sector 120
A.6 Example of Two Industries in Uberlandia Before and After Road Investment 121
A.7 Variation in Local Travel Times in Uberlandia, 2007–2014 122
A.8 Distribution of Change in Co-Location 123
A.9 Average Internal and External Agglomeration Across Industry Pairs. Count- and Employment-Based Measures 125
A.10 Average Distance Across Industry Pairs. 125
A.11 Location of All Stand-Alone Firms in Plastics Materials (SIC 282) and Petroleum Refining (SIC 291). 126
A.12 Location of Plants in Plastics Materials (SIC 282) and Petroleum Refining (SIC 291) of one Multi-Business Firm. 127
A.13 Share of Multibusiness Firms in all Firms, Establishments, and Employment, Brazil 2004–2014. 128

Acknowledgements

First, I would like to acknowledge the encouragement and support of my advisors, without whom this work would not have been possible. My sincere gratitude goes to Juan Alcácer who never stopped short of pushing the boundaries and motivating me to aspire to the highest standards of academic work. I thank Laura Alfaro for her mentorship, for supporting various aspects of my development, and for sharing generously her time and contagious enthusiasm every step of the way. I thank William Kerr for his excellent advice, guidance, and steadfast support. Beyond grateful, I am inspired by the example each of you set in your work and scholarship.

I also wish to express my gratitude to Richard Zeckhauser who thought me many essential maxims to guide both life and work. To Eric Werker and Jeffry Frieden for their initial support of this pursuit. A special thank you to my excellent teachers Pol Antràs, Mark Melitz, and Jordan Siegel for inspiring many aspects of my research. To my Center for Internal Development (CID) family, especially Ricardo Hausmann, Frank Neffke, Ljubica Nedelkoska, Michele Coscia, and Sebastián Bustos — thank you for allowing me to keep a home on both sides of the river.

To my cohort of doctoral students and in particular Kirill Borusyak, Yo-Jud Cheng, Cheng Gao, Maria Ibanez, Do Yoon Kim, Hyunjin Kim, Christopher Poliquin, Haris Tabaković, Mike Teodorescu and members of the strategy seminar series — thank you for many challenging and inspiring conversations. I am also grateful to the all-around support from the doctoral programs office at Harvard Business School and the financial support received for this work from the Harvard University David Rockefeller Center for Latin American Studies and the Harvard University Department of Economics Pellegrini Research Grant.

In this work, I benefitted from the unconditional support of my family — my mother Zorica and brother Edin, my parents-in-law in law Noshy and Rigo, my children Sofia and Adrian — and our Cambridge "village" that it took. Last but not least my gratitude to JP, for endless encouragement.

To my mother.

Introduction

Where a firm chooses to locate is a key strategic choice, one that affects many subsequent choices. For example, a firm's location affects which customers it serves, what kinds of employees it can attract, which inputs and resources it can access, who it competes with. From a societal perspective, where firms — and in particular large, multibusiness, multi-plant firms — choose to locate can affect the economic opportunities and subsequent development of regions as whole.

In this dissertation, I study the drivers and effects of firms' location choices, focusing in particular, on differences in the behavior of simple single-unit firms and large, complex, multibusiness and multi-plant firms. Beyond studies of multinational firms in the international context, research of firm location and economic geography has tended to abstract from firms' organizational form. Thus, the choice of a location has often been modelled as a firm choosing over a set of features in the *external* environment. This dissertation, on the other hand, incorporates the idea that multi-plant firms face a more complex decision, as they optimize location over their various units, taking into account *internal* interdependencies, as well as the external environment. Thus, rather than a single location choice, they optimize over a broader "spatial strategy."

Under this agenda, the first chapter of the dissertation focuses on uncovering the value of one aspect of a location — the proximity to other firms in one's industry. The focus of this Chapter is on stand-alone firms. In theory, proximity to agglomerations and clusters of economic activity can generate benefits for firms by allowing them to access broader pools of specialized inputs and resources. However, the empirical evidence regarding the benefits of co-location is mixed. In this paper, I seek to gain a better understanding into the source of these conflicting findings. I propose one explanation: heterogeneous effects of proximity across different industries. While in some industries, positive spillovers between firms may be important and large, in other industries proximity can be a net negative force by increasing competition.

To test this argument, I study how firm survival is affected when firms are brought closer together, not through choice, but through a change in the external environment — a decrease in travel times resulting from improved roads. The study is set in a large emerging market, Brazil, where I quantify how a seven-year government road investment program affected the *effective* proximity between firms. My results show that increased proximity in industries that compete locally leads to higher exit rates among the smallest firms and increased survival rates of the largest firms, effects which are consistent with increased competition. In contrast, in industries that do not compete locally, increased proximity improves the survival prospects of firms of all sizes. Beyond producing evidence for the effects of co-location, this study also illustrates how infrastructure investments, a top priority for many governments around the world, can change the benefits of locations and affect the pre-existing patterns of performance among incumbent firms.

Chapter 2 of the dissertation turns to the question of location strategy in multibusiness firms. Guided by the view that both internal and external factors affect their location choice, we study to what extent multibusiness firms exhibit "internal agglomeration" — greater co-location than would be expected if multibusiness firms were choosing locations like standalone firms do, without regard for internal proximity. Using data on a large sample of U.S. manufacturing plants and industries and well-established measures of agglomeration, we calculate agglomeration within the firm (internal) and agglomeration between firms (external) separately. We find that internal agglomeration is large: at any distance threshold, plants co-locate inside firm boundaries at about twice the rate that standalone firms co-locate. We next investigate whether internal agglomeration is driven by the desire of multibusiness firms to share resources among their plants. We find that greater resource relatedness between industries is predictive of greater agglomeration and, in particular, that the potential for labor pooling exerts the most significant force in keeping multibusiness firms' plants closer together.

Intrigued by the findings of the importance of labor market pooling in multibusiness firms, Chapter 3 investigates internal labor markets in multibusiness firms directly. Existing literature has tended to focus on "vertical" internal labor markets, the ways in which workers to move up the firm's hierarchy over their careers. Meanwhile we investigate the "horizontal" function of internal labor markers in multibusiness firms, which have received less attention. We seek to distinguish a hypothesis that internal labor markets exist to avoid external labor market frictions (e.g. high costs of hiring and firing) from a hypothesis that internal labor markets optimally allocate firm-specific and valuable knowledge embedded in workers to the different units of the firm.

We find that the redeployment of workers inside multibusiness firms is frequent, with an average of 12 percent of new workers hired in any year coming from within the firm. We find that mangers, in particular, have a much higher probability of being redeployed compared to other types of workers. More firm-specific experience is a significant predictor of more redeployment. We find that redeployed workers earn a 9 percent higher salary than similar workers simultaneously hired into the same position and plant via the external labor market, and this premium is higher, the higher their firm-specific experience. We also document that geographic distance between a firm's plants strongly discourages redeployment, findings that are consistent with the internal agglomeration results of Chapter 2. Overall, this chapter suggests that internal labor markets are an important and valuable tool for optimally allocating resources in multibusiness firms.

The overarching goal of the three chapters of the dissertation is to increase our understanding of both the optimal and actual patterns of firm location choices, especially to deepen the investigation into the more complex spatial strategies in multibusiness firms. If we can better understand the tradeoffs that large firms make as they seek to expand geographically, we may be better able to advise companies on their location strategies and well as to inform policymakers who seek to attract and create the right ecosystems for such companies to thrive in their regions.

Chapter 1

When Distance Shrinks: The Effects of Competitor Proximity on Firm Survival

1.1 Introduction

Geographic proximity between firms in the same industry, co-location, is a double-edged sword. In theory, it can give rise to positive agglomeration economies — benefits that stem from denser local pools of workers, inputs, ideas, or demand (Marshall 1920). But co-location can also breed competition over local customers (Hotelling 1929) and resources (Hannan and Freeman 1977). The empirical literature asking how co-location of firms in the same industry affects performance has arrived at positive, negative, and null results depending on the set of industries and firms analyzed (e.g. Baum and Mezias 1992, Sorenson and Audia 2000, Vernon Henderson 2003, Beaudry and Swann 2009, Buenstorf and Klepper 2009). This state of the literature, which has been referred to as “tentative" (McCann and Folta 2008) and “troublesome" (Arikan and Schilling 2011), might not be entirely surprising in light of theory. It points to the possibility that the net effect of agglomeration benefits and competition forces differs across firms and industries. The question, then, is: which firms benefit and which are hurt by proximity to competitors?

In this paper, I provide causal evidence of a key dimension of heterogeneity that determines how co-location affects firm performance: the geographic scope of competition. Combining insights from models of spatial competition and of agglomeration (Hotelling

1929, Syverson 2004, Combes, Duranton, Gobillon, Puga and Roux 2012), I argue that in industries where firms compete locally (e.g. concrete, dairy), the competition effects dominate and increased proximity between firms leads to a process of competition-driven selection, i.e. a “weeding out" of the least productive firms from the market. Meanwhile in nationally traded industries (e.g. car parts, electronics), the agglomeration benefits dominate and increased proximity improves the survival prospects of co-located firms.

To identify the effects of co-locating with competitors on firm performance I introduce a new empirical approach, which uses road upgrades as an exogenous shock to co-location. It addresses problems associated with endogenous firm location choice and unobserved heterogeneity that plague existing measures based on firm entry, exit, and growth dynamics. It also echoes recent findings that actual costs of mobility and communication like those shaped by the road network, rather than distance per se, define cluster shapes (Kerr and Kominers 2015) and the spatial scope of competition (Haveman and Rider 2014).

The setting is Brazil, where the federal government invested more than 70 billion reals (roughly US$ 35 billion) during 2007–2014 in road upgrades under the Programa de Aceleração do Crescimento (“PAC"). Building on methods from the burgeoning literature in spatial economics, I collect detailed geospatial data on the road network and its condition before and after the program to calculate how it affected travel times between all Brazilian municipalities (geographic units slightly smaller than U.S. counties). I combine the travel time data with exhaustive data on the location and operations of all formal sector manufacturing firms to measure how co-location changed only due to changes in minimum travel times stemming from road upgrades. In order to avoid well-known problems in using arbitrary administrative boundaries, I define each firm’s local market organically, as the geographic area that it can reach within four hours (a one day’s drive including return trip).

An important concern is that road upgrades may be targeted to particular locations or industries in ways that are endogenous to their expected performance. My identification strategy addresses this possibility by including a rich set of fixed effects for each Brazilian municipality and each industry. These control for any overall correlation between the road investments and the subsequent performance patterns of firms in a region or industry. The remaining variation, then, that the paper exploits is at the municipality-industry level — i.e. differences in how much road upgrades affected different industries in the same municipality. This variation stems only from differences in industries' pre-existing location patterns. The intuition behind the identification strategy can be illustrated with a simple example. Consider two firms, one producing ice cream and the other soft drinks in the Brazilian municipality of Uberlandia. Assume that in 2007, other ice cream producers in the vicinity of Uberlandia happened to lie mostly to the west while other local soft drinks manufacturers were located mostly to the east of Uberlandia. If Uberlandia saw an upgrade on a road leading westward, the ice cream producer would see a larger shock to co-location than the soft drink manufacturer. Unless the government program systematically targeted specific industry-municipality pairs in the manufacturing sector,[1] this strategy will yield causal identification of the effects of a change in co-location on performance.

Studying the effects of changes in co-location on manufacturing firms' survival rates, I find strong support for the prediction of heterogeneous effects in locally and nationally traded industries. Using the Ellison and Glaeser (1997) index to identify which industries geographically spread out (locally traded) versus concentrate (nationally traded), I find that in locally traded industries, doubling co-location reduces the survival probability of the smallest firms by 14.1 percentage points and increases the survival rate of the largest firms by 2.6 percentage points (given an average seven-year survival probability of 46.2 and 64.1 percent, respectively). This firm-level heterogeneity is in line with the prediction of competition-driven selection, i.e. the weeding out of the smallest firms from the market, and reallocation of market power to the largest firms. In contrast, in nationally traded industries doubling co-location increases firms' survival probability by 14.9 percentage points (given

[1]Conversations with the government secretariat in charge of the PAC road investment program suggest that initial project choice was determined to a large extent by the existence of feasibility studies (which often take several years to prepare) and capacity constraints on the existing network. Another objective was to distribute investments geographically, so that all states received some investment. To the extent that economic considerations entered investment decisions in later years, those tended to be sensitive to the needs of largest exporting sectors (e.g. soy and corn) rather than manufacturing. Appendix Figure A.2 illustrates the extent and locations of the road investments programmed under PAC.

an average seven-year survival probability of 55.4 percent), with no significant differences across firm sizes.

Additional analyses that study firms' strategic responses to the road shocks, specifically the prevalence of product-switches and relocations, find further evidence consistent with the heterogeneous effects hypotheses. I find that when co-location increases, firms in locally traded industries become more likely to switch their primary product or relocate to a different municipality. Firms that switch or relocate, tend to do so in a way that *decreases* co-location relative to their original choice of industry and location, a finding suggestive of competitive repositioning (Wang and Shaver 2013). Meanwhile in nationally traded industries, when co-location increases, firms become less likely to relocate. When they relocate or switch their primary product, they tend to *increase* co-location by moving toward firms in their industry, a finding that is consistent with resource-seeking agglomeration (Chung and Alcácer 2002, Kalnins and Chung 2004).

I consider a number of alternative explanations and find that these results are not driven by increased local market access, competition from new entrants, importing and exporting firms, or pre-period trends. I also do not find evidence of surviving companies getting bigger, e.g. due to consolidation or effects on the labor market via higher wages. While I cannot test the mechanisms directly (without observing local prices, knowledge flows, etc.), the results are consistent with proximity increasing competition in locally traded industries and facilitating agglomeration spillovers in nationally traded industries.

This paper contributes to the empirical literature on the performance effects of co-location. It identifies a key dimension of heterogeneity in the proximity-performance relationship across industries which can help to explain why studies that focus on single industries may arrive at positive (Chung and Kalnins 2001, Henderson 2003), negative (Baum and Mezias 1992, Sorenson and Audia 2000), mixed (Beaudry and Swann 2009), and null results (Buenstorf and Klepper 2009) while those that estimate average effects across manufacturing may find small (Martin, Mayer and Mayneris 2011) or null effects (Basile, Pittiglio and Reganati 2017). Beyond industry-level heterogeneity, I also find differences in the effects of co-location for small and large firms. Methodologically, I introduce a new empirical approach, leveraging changes in actual travel times to arrive at alternative, more exogenous, measures of changes in co-location.

This paper also contributes to an active empirical literature on the effects of transportation infrastructure investments. While governments around the world spend billions of dollars annually on transportation infrastructure, until recently, causal evidence regarding the impact of such investments was lacking. New advances in digital mapping and geo-spatial data analysis have given rise to a flurry of research in this area. We now have evidence that infrastructure investments affect welfare (Donaldson 2018), regional growth dynamics (Faber 2014, Goswami, Ghani and Kerr 2016), innovation (Agrawal, Galasso and Oettl 2017), and new firm entry (Gibbons, Lyytikäinen, Overman and Sanchis-Guarner 2017). This paper highlights the effects of such investments on competition and agglomeration spillovers, revealing important heterogeneity in how firms are affected by increased proximity to other firms in their industry.

This paper also relates to research on the effects of geographic distance in international settings (e.g. Nachum and Zaheer 2005, Ghemawat 2007, Boeh and Beamish 2012) and the effects of increased international integration (e.g. via lower trade barriers or foreign direct investment) on agglomeration and competition in local markets (Aitken and Harrison 1999, Ghemawat and Thomas 2008, Martin, Salomon and Wu 2010, Alfaro and Chen 2017). The finding that the effects of lowering mobility barriers *within* national borders produces similar effects to those *across* borders, invites further inquiry into the somewhat overlooked role of local distances in firms' strategic interactions.

1.2 Conceptual Framework

The key argument that I put forth in this section is that the effect of increased co-location will differ for firms in industries that compete locally and those that compete nationally. First, I summarize the state of the literature which motivates an inquiry into industry-level differences. I then discuss conceptually the nature of locally and nationally traded industries and sketch a framework to highlight the source of heterogeneous effects. The framework informs the hypotheses tested in the empirical part of the paper.

1.2.1 Co-location, Competition, Agglomeration

Strategy research has acknowledged the duality of proximity as a force of both competition and agglomeration early on and has asked how these dual forces shape firm location strategy. In one of the earliest studies to do so, Baum and Haveman (1997) show that product and geographic location choices of newly founded hotels in Manhattan appear to balance agglomeration (co-locating with similarly priced hotels) with differentiation on size. Subsequent studies show further heterogeneity in how firms trade off agglomeration benefits and competition in their location choices. For example, Shaver and Flyer (2000) show that small firms co-locate (seek agglomeration benefits) while large firms no not (avoid competition). Alcácer (2006) shows that firms co-locate versus disperse depending on the value-chain function, seeking spillovers in R&D activities but avoiding competition in sales.

While the literature on location choice illustrates strong patterns of co-location (and dispersion), studies on the subsequent link between the degree of firm co-location and performance are far less conclusive. Looking at the manufacturing sector as a whole, evidence tends to favor small positive effects of co-location on productivity as measured by sales or total factor productivity (e.g. Chung and Kalnins 2001, Henderson 2003, Martin, Mayer and Mayneris 2011). However, studies of the effect of co-location on firm survival have arrived at positive (Pe'er, Vertinsky and Keil 2016), negative (e.g. Baum and Mezias 1992, Sorenson and Audia 2000, Folta, Cooper and Baik 2006), mixed (Beaudry and Swann 2009), and null results (Buenstorf and Klepper 2009) depending on the context and industry analyzed.

In a review of state of the literature, McCann and Folta (2008) refer to all conclusions on the proximity-performance relationship as "tentative", an assessment, which is echoed in a more recent review (Frenken, Cefis and Stam 2014), with both studies agreeing on the need of future research to reconcile contradictory empirical findings. The lack of clear empirical findings calls for a focus on the sources of heterogeneity to provide further insights. In this spirit, recent papers have looked to firm-level factors (including firm age, knowledge stock, size, foreign or domestic status) to ask whether some firms benefit from co-locating more than others (e.g. McCann and Folta 2011, Rigby and Brown 2015).

In contrast to firm-level heterogeneity, industry level heterogeneity has been largely unexplored (beyond a generic split of manufacturing versus services), although it may be important and could explain inconsistent findings across different studies. Specifically, if the effects of agglomeration are heterogeneous across industries, then single-industry studies, while informative of the balance of competition and agglomeration in a given industry, might not be generalizable to other industries. Meanwhile, studies of the effects of agglomeration in the manufacturing sector as a whole or in broad sub-sectors may be averaging positive effects in some manufacturing industries with negative effects in others.

1.2.2 A Framework with Heterogeneous Effects

To sketch a framework of industry-level heterogeneity, I leverage insights from two research streams, one rooted in industrial organization (IO) and the other in urban economics research, which provide theoretical grounding for questions on the proximity-performance relationship. IO models of spatial competition (Hotelling 1929, Salop 1979, Syverson 2004) focus on the competitive dynamics between co-located firms. A key idea in these models is that the number and proximity of competing firms provides downward pressure on prices and markups, thus squeezing profits. Agglomeration theories, on the other hand, focus on the potential productivity enhancements that proximity between firms can provide, in particular the positive externalities arising from shared pools of local input suppliers, specialized workers, and knowledge (Marshall 1920).[2]

The incorporation of both competition and agglomeration mechanisms into standard economic models is relatively recent (e.g. Combes, Duranton, Gobillon, Puga and Roux 2012)

[2]Agglomeration theories also acknowledge the potential of agglomeration to heighten competition for resources on the supply side, increasing local wages or rents. However, most empirical work focuses on the net productivity benefits, which are expected to outweigh the potential costs.

and existing empirical tests of their relative importance are few. The question remains: if co-location affects firms both via agglomeration spillovers and competition, what determines their relative importance? I argue that the geographic scope of competition — whether firms compete for customers in the local market or on broader national or international markets is key in determining the nature of the relationship between co-location and firm survival.

To illustrate the theoretical mechanisms behind the prediction of heterogeneous effects, next I sketch a simplified framework, which introduces an industry-level parameter that determines the relative weight of competition and agglomeration spillovers in a firm's profit function. Consider a representative firm i in industry s and municipality m earning profits that are a function of revenues minus the costs of labor and capital incurred in production:

$$\pi_i^{sm} = P_i \times Q_i - wL_i - iK_i$$

I assume that co-location with firms in its industry will affect the firm's profit function in two separate ways: i) by enhancing its productivity on the supply side (increasing Q_i) and ii) by increasing price competition on the demand side (lowering P_i).[3]

Following standard agglomeration models, the supply side productivity increase due to agglomeration is modeled as an increase in the firm's total factor productivity (TFP). Specifically, consider that output is a function of TFP and inputs: $Q_i = A_i K_i^{\alpha} L_i^{\beta}$. TFP itself is an increasing function of a firm-specific productivity parameter ψ_i and co-location:

$$A_i = f_1(\varphi_i) \ (CoLoc^{sm})^{\gamma_s}$$

The productivity parameter can be thought of as the firm's random draw from a productivity distribution in the spirit of Lippman and Rumelt (1982) and Melitz (2003). Greater co-location in this framework enhances TFP by facilitating positive spillovers in inputs, workers, and

[3]Note that for simplicity, I assume that in the short- to medium-term, co-location does not affect the prices of inputs, i.e. the wage rate and the rental rate of capital in the framework above. Empirically, this amounts to assuming that changes in co-location due to the roads did not change wages and interest rates in the 2007–2014 period. I check for any wage effects in the empirical exercise and indeed, find no evidence of changes in wages for incumbent firms. To the extent that prices of inputs were affected, any positive effects of increased co-location would be outweighed by higher input prices, making any positive agglomeration effects more difficult to detect.

knowledge between firms. Importantly, the extent to which it does so differs by industry. This industry-level heterogeneity is incorporated into the framework with the parameter $\gamma_s \in [0,1]$.

Why would the effect of co-location on productivity vary by industry? Industries differ in how much they rely on specialized inputs, workers, and knowledge. While high-end products heavily rely on such specialized resources, they tend to be less important in industries which require few inputs, or where inputs are relatively homogenous. Note that in the extreme case in which $\gamma_s = 0$, a firm's TFP depends only on its own productivity draw and not at all on co-location.

Next let's consider how co-location may affect competition on the demand side. Models that allow for strategic interactions on the demand side have the key feature that a firm's price is sensitive to those of competitors — i.e. prices are determined through strategic interaction. A general feature of spatial competition models is that prices are decreasing in the firm's own productivity in the productivity of other firms in the local market — i.e. co-location:

$$P_i = f_2(\frac{1}{\varphi_i}) \left(\frac{1}{CoLoc^{sm}} \right)^{\rho_s}.$$

An extreme case of such models are spatial competition models where competition is extremely local — between neighboring firms. Support for such very localized competition exists for some industries, for example in the wholesale gasoline market (e.g. Pinkse, Slade and Brett 2002). One can imagine differences across industries in the degree to which competition is localized, which is summarized in the parameter $\rho_s \in [0,1]$. While in some industries a firm's price is sensitive to those of its nearest neighbors, in others, prices are determined nationally, or in the extreme case, demand is elastic at an exogenous price $\bar{P}$.

1.2.2.1 Comparative Statics

Revisiting the profit equation, having described how co-location theoretically enters into prices and quantity produced, one can see that an exogenous increase in co-location will potentially have two opposing effects — increasing the quantity that a firm can produce with a given set of inputs through a TFP boost but also lowering prices due to competitive price pressure. Which effect is larger, and therefore, what the net effect of an increase in co-location on profits is, will depend on the relative sensitivity of productivity and prices to co-location, i.e. the particular values of the industry parameters γ_s and ρ_s. For illustration, it is useful to consider two extreme cases and an intermediate case:

Case 1: $\gamma = 0$ and $\rho = 1$. This is a case when a firm's productivity is not at all responsive to co-location via agglomeration economies but its price function is highly sensitive to the presence of local competitors. Examples are industries utilizing few specialized inputs, labor, or knowledge but which compete locally, for example water bottlers. In this case, co-location only effects profits though the competition channel and the effect is negative, with higher levels of co-location leading to lower profits. More formally:

$$\frac{\partial A}{\partial CoLoc} = 0, \; \frac{\partial P}{\partial CoLoc} < 0 \Rightarrow \frac{\partial \pi}{\partial CoLoc} < 0$$

Case 2: $\gamma = 1$ and $\rho = 0$. This is a case when productivity is very sensitive to co-location via agglomeration economies but price is not. Examples are industries that rely on highly specialized inputs, workers, and knowledge but for which prices are determined in the national market, for example manufacturers of consumer electronics. In this case, co-location only affects profits through higher productivity and the effect is positive, where higher levels of co-location lead to higher profits. More formally:

$$\frac{\partial A}{\partial CoLoc} > 0, \; \frac{\partial P}{\partial CoLoc} = 0 \Rightarrow \frac{\partial \pi}{\partial CoLoc} > 0$$

Case 3: $\gamma > 0$ and $\rho > 0$. In these intermediate cases, both agglomeration economies and competition effects enter the profit function and the overall effect of increased co-location depends on the relative size of the two effects:

$$\frac{\partial A}{\partial CoLoc} > 0, \; \frac{\partial P}{\partial CoLoc} < 0 \Rightarrow \frac{\partial \pi}{\partial CoLoc} > 0 \, if \; \gamma_s \gg \rho_s \; and \; \frac{\partial \pi}{\partial CoLoc} < 0 \, if \; \gamma_s \ll \rho_s$$

Based on the arguments above, I expect that in industries which compete locally, prices are sensitive to the presence of local competitors, and that greater co-location will translate into lower prices. Moreover, these industries tend to be bulky, undifferentiated goods where agglomeration economies are less likely to be relevant. On the other hand, in industries that compete nationally, the change in prices in response to co-location in a particular market is likely to be negligible. These tend to be more high-end, products or resource-based products requiring specialized inputs or distribution channels — here agglomeration benefits are more likely to be positive.

In this framework, how will an exogenous increase in co-location affect the survival prospects of an incumbent firm? A key insight of heterogeneous firm frameworks (e.g. Melitz 2003, Syverson 2004, Combes, Duranton, Gobillon, Puga and Roux 2012) is that, given a distribution of firm productivity, there exists a cutoff productivity level below which it is no longer profitable for a firm to continue production but rather to exit. Denoting the original level of this threshold with ψ^0, a change in co-location which lowers the price charged by producers in the market will make some firms that were just making non-negative profits now no longer feasible, forcing them to exit. Intuitively, the cutoff level of productivity above which firms participate in the market will increase, forcing firms at the bottom left portion of the firm productivity distribution to exit (Figure 1.1, left). Their market is re-allocated to the more productive surviving firms, which in addition to the price decrease, may now see an increase in market size. While these forces go in opposite directions, it is likely that following more exit of less productive firms, the survival prospects of the most productive firms will increase due to the market size effect. These expected dynamics in locally traded industries lead to the following hypothesis:

Hypothesis 1: In locally traded industries, an increase in co-location will lead to increased exit of the least productive firms.

Hypothesis 2: In locally traded industries, an increase in co-location will improve the survival prospects of the most productive firms.

Meanwhile, a change in co-location which increases the productivity of all local firms through higher TFP but is not reflected in product prices (which may be determined by a much larger number of firms beyond the local market), has the effect of facilitating the

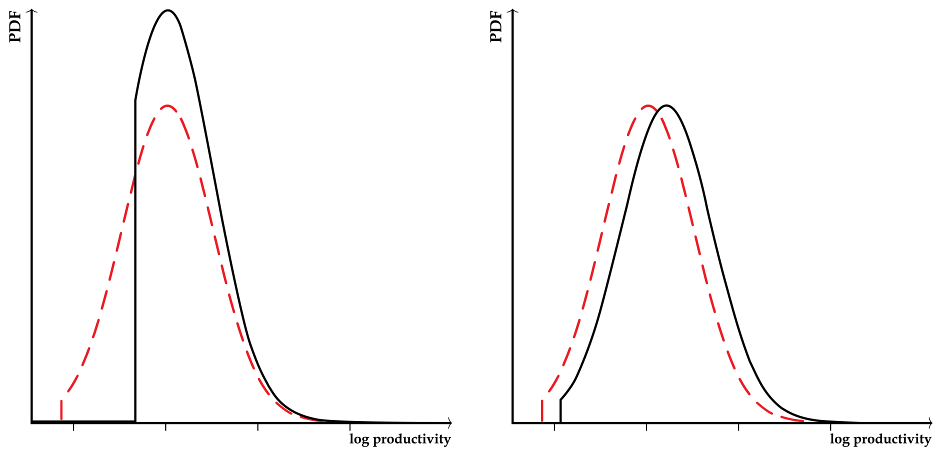

Figure 1.1: *Predicted Effects of Increases in Co-Location in Locally Traded (Left) and Nationally Traded (Right) Industries*

Note: The Figure represents a stylized example of the hypothesized distribution of log firm productivity before (dashed line) and after (solid line) an increase in co-location. The probability density function (PDF) represents the number of firms with each level of productivity. Reproduction following Combes et al. (2012), Alfaro and Chen (2017).

survival of all firms in the local market. Visually, it can be represented as a shift in the entire distribution of firm productivities to the right (Figure 1.1, right). These dynamics in nationally traded industries leads to the following two hypotheses:

Hypothesis 3: In nationally traded industries, an increase in co-location will improve firms' survival prospects.

1.3 Data and Measurement

1.3.1 Measuring Co-Location

Two standard ways to measure co-location are used in the literature. Both count the number of firms (or workers) located the same industry as the focal firm and pre-defined geographic area. Differences arise in the choice of geographic area. While discrete measures count all firms in the same administrative unit (e.g. state or county), continuous measures use spatial distance bands (e.g. 10km, 50km) without regard for administrative boundaries, normalizing the count by the distance to the focal firm (Sorenson and Audia 2000, Rosenthal and Strange 2003, Haveman and Rider 2014). The second approach is more representative of real-world interactions, which do not make a full stop at administrative boundaries. There is no consensus in the literature on what the "right" geographic distance over which to generate the count is; rather it depends on the economic interactions that the researcher is interested in (Alcácer and Zhao 2016).

In this paper, I follow the second approach but with two important refinements. First, I use a geographic definition of market that is independent of administrative boundaries, defining local markets "organically" as all of the destinations that a firm can reach in four hours traveling on local roads. I choose four hours because of its practical significance as the destinations that can be reached with one working day (including the return trip), keeping in mind relatively strict regulations regarding 8-hour working days for truckers in Brazil. Appendix Figure A.1 shows examples of markets arising from this definition. Second, rather than depreciating the count of each firm by its straight-line distance to the focal firm, I depreciate by the actual travel time, thus generating a measure that better incorporates the travel possibilities and costs between locations. Specifically,

$$CoLoc_{it}^{sm} = ln\left(\sum_{k\in M}\frac{x_{jt}^{sk}}{\tau_{mk}}\right)$$

is the co-location of firm i at time t in industry s and municipality m, x_{jt} is the variable being counted (firm j in industry s and municipality k at time t), τ_{mk} is the travel time between municipalities m and k, and M is firm i's local market defined as all of the destinations which are no more than four hours apart from municipality m by road, i.e. the set of all k such that $\tau_{mk} \leq$ 240 minutes.[4] Because a small number of firms have zero other firms in their industry within their four-hour radius, I add 0.1 to the raw measure before taking logs

[4]Because I don't observe within-municipality travel times, for any pair of firms that are located in the same municipality, I set the within-municipality travel time equal to 15 minutes. The average size of a municipality in Brazil is roughly 1,500 square kilometers. Assuming municipalities take the shapes of a square, one can calculate the expected distance between two randomly chosen points to be roughly 20 kilometers, which translates into 15 minutes driving time at a speed of 80 kilometers per hour.

in order to preserve these observations.

1.3.2 Firm-Level Data

Firm level data come from the *Relação Anual de Informações Sociais* (RAIS) a high-quality employer-employee matched database that provides an annual census of all formal-sector firms in Brazil. I aggregate the employee-level data to the establishment level using each establishment's unique ID to create an establishment-level panel dataset.[5] I restrict the sample to firms active in manufacturing during the entire period with a minimum size of three workers over the 2007–2014 period.[6] I drop industries which may be directly affected by road construction, industries with fewer than ten firms active in the baseline year 2007, and three industry categories that are too general to be meaningful for identifying competition or agglomeration forces.[7] In the analyses of firm performance outcomes (but not in the measure of co-location), I also exclude all companies with state ownership or control to lessen concerns around special treatment of such companies and establishments belonging to multi-unit. For multi-unit firms, changes in travel times may incentivize plant exit but due to other mechanisms, e.g. relating to their internal organization and internal agglomeration (Alcácer and Delgado 2016).[8] Given this exclusion, all firms in the baseline

[5]When aggregating the employee-level data to the establishment level, an establishment can become associated with more than one record in the same year due to differences in the employee-level data in field values for industry, legal form, municipality, etc. In such cases, I select the modal sample, meaning the record associated with the largest number of employee contracts. I further clean the data by removing establishments with invalid IDs, all CEI entities which are multi-jurisdiction entities primarily associated with the construction sector, and all organizations that were not a business entity during the entire period.

[6]Firms in manufacturing industries are identified using information on the industry code reported by each establishment. In the Brazilian industry classification system, Classificação Nacional de Atividades Econômicas (CNAE) version 1.0, manufacturing industries fall in 2-digit sectors 15-37. For more detail, see http://biblioteca.ibge.gov.br/visualizacao/livros/liv2314.pdf.

[7]The seven industries excluded due to the first criterion are manufacturing of concrete (2630), stone (2691), cement (2691), construction bulldozers (2953), construction equipment (2995), earthmoving and paving equipment (2954) and heavy military equipment manufacturing (2972). The three excluded due to the second criterion are industries relating to chemical products, machine repair, and motor vehicle parts containing "not elsewhere specified" in the industry name.

[8]Multi-firms make up around seven percent of manufacturing establishments in Brazil. Note that both state-owned and multi-unit firms' establishments are included in the calculations of all economic variables, including co-location, as these firms do affect the local competitive and agglomerative dynamics.

sample are single-establishment firms.

The analysis will examine the survival dynamics of the firms in this sample from 2007–2014.[9] Looking at 2007 incumbents, i.e. all firms in business in 2007 that entered in 2006 or earlier, this sample consist of close to 130,000 firms. Appendix Table **??** shows their descriptive statistics. The average firm is xx years old and employs xx full-time-equivalent workers. Overall, the firms in the sample represent 252 unique manufacturing industries and 3,724 unique Brazilian municipalities.

1.3.3 Travel Time Data

I combine three data sources to construct the time-varying travel times between all Brazilian municipalities in 2007 and 2014. The first is data on the location of municipal capitals and on the federal and state road network, which come from the Brazilian Ministry of Transportation National Logistics and Transport Plan (PNLT) maps.[10] The PNLT map details the spatial location, length, and surface type of the entire Brazilian federal and state road network as of 2009, which consists of more than 18,000 distinct road segments with a total length of more than 280,000 kilometers.

The second data source comes from National Department of Transportation Infrastructure (DNIT), a federal government agency that oversees road infrastructure investments, which publishes data showing the surface type (e.g. duplicated, single-lane, dirt, etc.) for of each road segment at every year-end. I match these data to the PNLT geo-spatial road network in order to construct a representation of the condition of the road network in 2007 (pre-PAC) and in 2014 (post PAC).[11] Appendix Table A.2 shows the percentage of the federal

[9] I exploit the 2000–2007 period of the data to control for pre-trends.

[10] The PNLT data were downloaded from http://www.transportes.gov.br/conteudo/2822-base-de-dados-georreferenciados-pnlt-2010.html. Prior studies that use this data include Schettini and Azzoni (2015) and de Carvalho et al. (2016).

[11] The DNIT data on road segment conditions are available from http://www.dnit.gov.br/sistema-nacional-de-viacao/. The match to PNTL data is made using the segment identifier (SNV/PNV código). Note that the segment identifier for some segments changes over time. Therefore, I first match all segments on the segment identifier and length. For the remainder, I apply a geo-spatial matching algorithm. The final translation file matching 2007 codes to 2014 codes is available from the author upon request.

road network in each surface type in 2007 and 2014.[12]

The third data source is a physical survey of all major Brazilian roads conducted annually by the Brazilian National Transport Confederation (CNT).[13] In 2007, the CNT surveyed more than 92,000 kilometers of the federal and state road network and assigned one of five indicators of surface condition (excellent, good, regular, bad, very bad). I digitized the 2007 and 2014 data and geo-referenced it to match the PNLT map.[14] Appendix Table A.3 shows the percentage of road segments by surface condition in 2007 and 2014. Overall, during the 2007–2014 period, the Brazilian road network saw t roughly 6,500 kilometers of road upgraded and roughly 25,000 kilometers of road improved — i.e. roughly 30 percent of the entire network saw some improvement.

Finally, with the information on road surface type and condition, I assign a travel velocity to each road segment. The velocity represents the likely actual speed of a truck driving on that road segment, rather than the speed limit. My assumptions of travel velocities for given road surfaces and conditions (Appendix Table A.4) are based on discussions with the Brazilian Ministry of Transport and leading Brazilian transport economists. The changes in travel velocity of the road segments, which stem from changes in the road surface type and/or condition between 2007 and 2014 are the only drivers of the changes in travel times that will be used in the analysis.

---

[12]Note that because data on the current surface is only available for federal road segments, I assume that the surfaces of state roads remain unchanged during the time period of analysis. This simplifying assumption is unlikely to create large distortions in the measures as during the 2007–2014 period, there were few major state-level road investment initiatives.

[13]Downloaded from http://pesquisarodovias.cnt.org.br/Edicoes in pdf format. I digitized and geo-referenced these data and matched them to the PNLT segments. These data are available from the author upon request.

[14]Over time, the CNT has increased the extent of the network that is surveys. In order to facilitate comparability, I use the CNT quality indicator only for those segments that were surveyed in both 2007 and 2014. All non-survey roads and those excluded based on the above criterion, are assigned a quality of "regular"for the purpose of the analysis.

1.3.4 Constructing an O-D Matrix of Travel Times

A key input into the analysis are changes in travel times between firm locations that result from the road improvements implemented during the PAC program between 2007 and 2014. The lowest geographic level at which I identify firm locations in the data are Brazilian municipalities, which are slightly smaller on average than U.S. counties. I identify each municipality as its main city (sede municipais) and use this as my definition of firm location.[15] Municipalities represent a finer level spatial unit of analysis than was available in most prior studies of the effects of roads (e.g. Faber 2014, Goswami, Ghani and Kerr 2016).

I overlay the location of all municipalities onto the geo-referenced road network to calculate the municipality-to-municipality minimum travel time on the road network in 2007 and in 2014. These origin-destination (O-D) travel times are calculated using the ArcGIS Network Analyst optimal routing tool. Taking each segment's length and velocity as an input, the program solves for the optimal path from each origin to each destination that minimizes travel time via Dijkstra's algorithm, a widely used algorithm for shortest path problems in graph theory.[16] The set of origins and destinations is formed by taking all municipal capitals that lie within 50 kilometers of a state or federal road.[17] The resulting O-D matrix is a roughly 30 million length matrix, containing the travel time (in minutes) and travel distance (in kilometers) between all city pairs in 2007 and 2014.[18]

---

[15]From time to time, new municipalities are created from one or more existing municipalities. In the 2000–2014 period, just over 50 "new" municipalities were created in Brazil. In order to ensure comparability of the concept of municipality over time, I follow prior studies in this context (e.g. Morten and Oliveira 2016) and construct "Minimum comparable areas" or MCAs which is the lowest geographic unit approximating a municipality that is constant since 2000. The 5,565 actual municipalities map to 5,478 MCAs. Throughout the analysis I refer to "municipality" and use the constant municipalities, i.e. the MCAs.

[16]Other recent papers have used similar techniques to estimate travel times between locations, for example Allen and Arkolakis (2014) and Morten and Oliveira (2016) who use the "fast marching algorithm". The FMA is similar to the Dijkstra shortest path algorithm, with the key difference being that the FMA can be applied to calculate speeds over continuous graphs (vs. networks) and can calculate the speeds of three-dimensional surfaces such as waves.

[17]Out of the 5,565 municipal capitals on Brazil, 5,505 meet this criterion. The majority of the municipalities that are not included due to this criterion lie in the Amazon region.

[18]Because some parts of the network are not connected to all other cities by road (i.e. some parts of the federal and state network form disconnected clusters), the actual length of the O-D matrix is 29.2 million.

Comparing the 2007 and 2014 O-D matrix that I derive, I find that *over the entire network, the program lowered travel times by roughly 5.2 percent*. This is a relatively large effect, compared to similar programs recently evaluated, for example in the U.K. (e.g. Gibbons, Lyytikäinen, Overman and Sanchis-Guarner 2017). Since the focus of this study is on changes in local markets, Figure 1.2 illustrates for each municipality the percentage change in the total travel time to all destinations that were reachable within four hours in 2007. While the average local travel time decrease is two percent, there is significant variation across different regions of the country, with some regions seeing travel times in local markets fall by more than 15 percent. Some regions experienced a net deterioration in the quality of local road networks, resulting in increases in local travel times.

1.3.5 Verifying O-D Matrix Accuracy via Google Maps

Since all of the variation used in the analysis stems from travel time changes, I implement several tests to verify the accuracy of the O-D matrix that the methodology yields. To do so, I select a one percent random sample of origin-destination pairs from the O-D matrix (292,040 city-pairs) and query each pair in this sample via Google Maps, recording the travel distances and travel times that it returns.[19] Despite some expected sources of differences between the estimates,[20] the exercise shows that the O-D estimates have a very high degree of overlap with the Google Maps results. Appendix Figure A.3 shows the correlation of the estimated travel distances (in kilometers) and travel times (in minutes) between the Google Maps results and the O-D matrix in the year 2014 which are 0.98 and 0.97, respectively. The

Specifically, 5,404 cities form the largest fully connected part of the network while the remaining 101 cities form smaller, disconnected clusters.

[19]The Google Maps queries were conducted in May 2017.

[20]There are several reasons why I do not expect the estimates to fully overlap with the Google Maps estimate. One is that the O-D matrix ignores the existence of local (municipal) roads, while these are taken into account in Google Maps. Second, Google Maps takes into account historical data on traffic congestion, while the O-D estimates do not. Both of these reasons are likely to lead Google Maps to predict longer travel times than the O-D matrix. There are also reasons that Google Maps would predict shorter times and distances and these include: i) The Google Maps query will incorporate any additional improvements in road conditions that took place between 2014 and 2017; ii) Google Maps will include the possibility of travel via ferry (i.e. not only via roads); and iii) Google Maps assumes velocities for light vehicles (cars) while my velocity assumption reflects the likely speed of trucks.

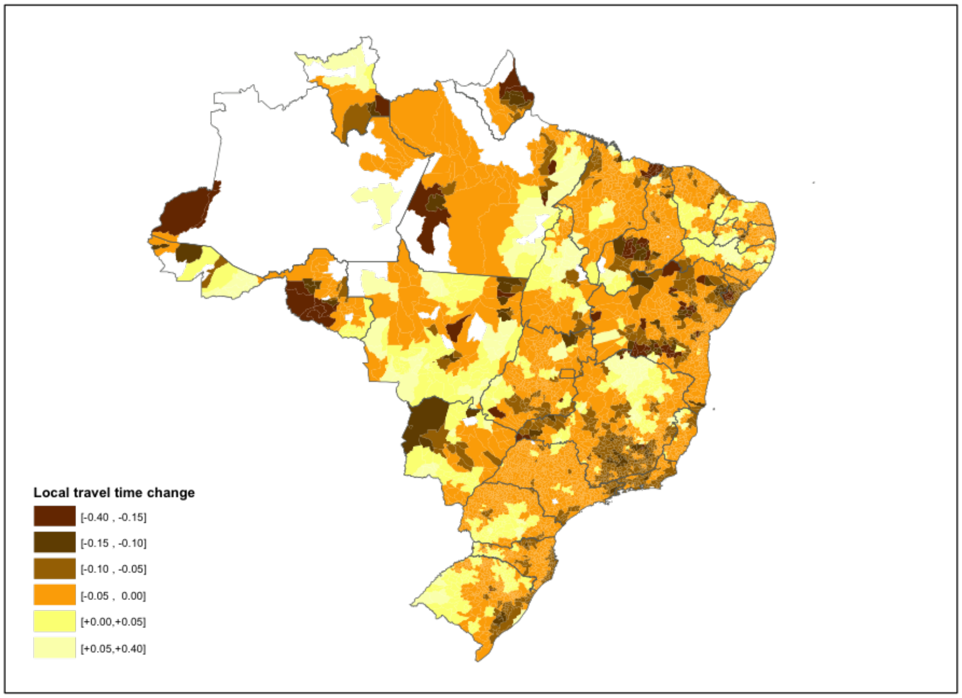

Figure 1.2: *Change in Local Travel Times Between in Brazilian Municipalities, 2007–2014*

Note: Own calculations based on data from the Ministry of Transport of Brazil. Map shows the total percentage change in the average time of traveling from each city (sede municipais) to each other city in its local market (a city's local market is any municipality that can be reached in four hours or less traveling by federal and state roads in 2007). Areas not shaded (white) are municipalities whose capital lies more than 50 kilometers away from a state or federal road and are not included in the analysis.

distributions of the two estimates also have a very high degree of overlap (Appendix Figure A.4).[21] Specifically, for city pairs lying within four hours from each other or less (which are the distances I employ in this analysis) the median discrepancy between the distance estimated by Google Maps and the O-D matrix is 1.5 kilometers and the median discrepancy in travel times, 11 minutes.

In further tests, I regress the residual from a regression of Google Maps travel time on Google Maps distance on the travel time estimated by the O-D Matrix. While the first

[21] Manual checks suggests that the tendency of the O-D measure to have a higher density at high values of distance relative to the Google Maps measure tends to relate to the possibility of travel on river routes, which are usually not included in the O-D analysis. This problem is less likely to be relevant in within local markets, which are the focus of this paper.

regression (Google time on Google distance) has a R^2 of 0.99, the results of this exercise show that the O-D matrix travel times have significant power to predict the residual variation. Specifically, a regression of the residual component on the O-D matrix travel time has an R^2 of 0.3 with a t-statistic of 127.7. This second tests also confirm that the O-D matrix measures are meaningful signals of true real-world travel times.

1.3.6 Locally and Nationally Traded Industries

The most direct way to identify which industries trade locally versus nationally would be to observe the actual patterns of internal trade in different industries. In the United States, the Commodity Flow Survey (CFS), a survey of the movement of goods collected every five years, provides information on the distances at which goods are shipped for a selection of industries in manufacturing, mining, wholesale, and selected retail and services establishments. Even at the very broad sector groupings provided in the publicly accessible data, we see large differences across sectors in how far goods are traded. For example, the median shipment distance for resource-based products, beverages, food, fabricated metals, wood products and paper products shipped on trucks is less than 200 miles, a radius of roughly two to four hours. Meanwhile the median distance that a product in machinery, computer products, electrical equipment and apparel is shipped by truck is more than 500 miles (Appendix Figure A.5).

In the absence of internal trade data at a detailed industry classification level, studies have classified manufacturing industries following the principle that nationally traded industries concentrate production in a few locations and ship their product to more distant markets while locally traded or non-traded industries are found everywhere (e.g. Delgado, Porter and Stern 2015, Mian and Sufi 2014). I follow this principle and use data on the actual industry location patterns in Brazil from RAIS to classify industries as nationally or locally traded. To do so I calculate, for each industry the Ellison-Glaeser index (Ellison and Glaeser 1997), a well-known measure of the spatial concentration of industries which counts the share of an industry's employment in a geographic unit and compares it to the share of the industry's employment nationally, while adjusting for the industry Herfindahl index to account for concentrations that are due to the industry simply having few firms.[22] Industries with an index value above zero exhibit grater geographic concentration than economic activity overall, while those with a value below zero concentrate less than expected if they followed the overall distribution of economic activity.

I employ the cutoffs used in Ellison and Glaeser to categorize industries as highly concentrated or not highly concentrated to classify industries as nationally or locally traded. Specifically, industries with index values above 0.05 and below 0.02 are classified as nationally and locally traded, respectively. This procedure yields 73 locally traded and 99 nationally traded industries. Looking at the largest industries in terms of the number of firms represented in the sample (Appendix Table A.5), the locally traded sample includes a number of relatively undifferentiated, transport cost intensive products (metal frames, ice cream manufacturing, bakeries, dairy, wooden articles, etc.) while the nationally traded sample features more differentiated industries (cosmetics, electronic materials, machine tools, etc.).

1.4 Identification Strategy and Institutional Context

The following section describes the existing challenges in the empirical research on the links between co-location and firm performance and the assumptions and potential advantages inherent in the use of road upgrades as an alternative method of causal identification in this context. I discuss the institutional setting and the methodology used to generate an alternative measure of changes in co-location.

[22] The geographic units employed in calculating the index are Brazilian regional urban divisions (Divisão Urbano Regional - Regioes Imediatas de Articulação Urbana), which are geographic units defined by the Brazilian statistical institute (IBGE) that are exhaustive of Brazil's land mass and centered around urban areas, comparable in spirit to U.S. economic areas. See https://ww2.ibge.gov.br/home/geociencias/geografia/default_divisao_urbano_regional.shtm.

1.4.1 Identification Strategy

The current empirical literature on the effects of co-location on firm performance has struggled to address two potential sources of bias. One is selection bias created by the endogenous sorting of firms to locations. If a priori higher quality firms select themselves into more (or less) competitive markets, we could see a positive (negative) relationship between co-location and firm performance even in the absence of any causal effects. We indeed have strong evidence that selection exists (e.g. Shaver and Flyer 2000, Kalnins and Chung 2004). One way to address selection bias is by using panel data (Baum and Mezias 1992, Sorenson and Audia 2000, Henderson 2003), which allows for the analysis of *changes* in firm level outcomes as a function of *changes* in the degree of co-location. Panel data can largely control for selection bias through the use of first-differenced or panel regressions with firm fixed effects.

However, a second concern has proven harder to address. Even with panel data, changes in co-location as measured by the researcher will reflect the dynamics of firm entry, exit, and growth. If unobserved local variables (e.g. location- or industry specific shocks) affect firm entry, exit, and growth as well as the performance of the focal firm, then the estimates of the effects of change in co-location measured this way on firm performance are biased. This second concern calls for the use of instruments or natural experiments (Combes, Duranton and Gobillon 2011). However, strong time-varying instruments for agglomeration, are difficult to find. One existing solution is to use of lagged levels of agglomeration as instruments for future changes (e.g. Martin, Mayer and Mayneris 2011). This instrument tends to be predictive but still suffers from endogeneity if the effects of agglomeration are dynamic, as we have reason to believe.

In this paper, I use changes in travel times between firm locations caused by road upgrades as an exogenous source of variation in co-location. The key advantage of this approach is that it is free of both selection bias and unobserved heterogeneity because it does not incorporate any information on the changes in the composition or growth of firms.

Specifically, I define change in co-location as:

$$\Delta\, CoLoc^{sm}_{i,07-14} = ln\left(\sum_{k\in M_{14}} \frac{x^{sk}_{j,07}}{\tau_{mk,14}}\right) - ln\left(\sum_{k\in M_{07}} \frac{x^{sk}_{j,07}}{\tau_{mk,07}}\right) \tag{1.1}$$

Note that the measure sums over only the incumbent firms in 2007 (including those which did not survive) and thus does not incorporate the endogenous entry, exit, and growth of firms in the market. The only variation in the change in co-location measure stems from i) the variation in travel times between 2007 incumbents, $\tau_{mk,14}$ and ii) any changes in the size of the local market M that are due to road upgrades. I take a difference of the logged values of 2014 and 2007 co-location, because this measure has the benefit of approximating percentage differences (for small enough changes such as here) while being easily interpretable in a regression.

The use of changes in travel times for exogenous variation in co-location is motivated by the view that the mechanisms underlying the effects of co-location — i.e. competitive dynamics and agglomeration spillovers — are in fact sensitive to actual costs and patterns of mobility rather than only distance per se. Indeed, we have recent evidence for the view that changes of the road network affect price competition (Asturias, García-Santana and Ramos 2015, Gross 2016), labor flows (Morten and Oliveira 2016), and knowledge flows (Agrawal, Galasso and Oettl 2017).

The key assumption made in order for the approach to be valid is that degree to which the roads affected the co-location of manufacturing firms with their competitors is exogenous to their expected future performance (conditional on the appropriate controls). The next section of this paper motivates this assumption by describing the institutional context and identification more carefully.

1.4.2 Institutional Context

The setting of this study is Brazil during the 2007–2014 period. This is an interesting setting for a number of reasons. One is its economic relevance as the world's 7th largest economy (IMF 2016) and the fourth largest recipient of FDI inflows among emerging economies

(UNCTAD 2016). The trend of a growing share of the world's economic activity shifting towards emerging markets calls a better understanding of the ways in which institutional differences across locations affect firm strategy and firm performance (Ricart *et al.* 2004, Khanna and Palepu. 2010). Second, the lack of quality infrastructure, which is a feature of many emerging markets due to their underdeveloped institutions (Henisz 2002), is especially pronounced in Brazil.[23] Brazil therefore offers an excellent setting for understanding to what extent a release of this constraint changes the nature of competitive interactions as well as the potential for positive externalities among firms.

This study leverages the Programa de Aceleração do Crescimento ("PAC"), a government investment program which took place during 2007–2014 and invested more than 70 billion reals (roughly 35 billion dollars).[24] Unlike road investment programs that took a holistic approach,[25] the Brazilian program was highly decentralized, with more than 250 different road investment projects taking place across different parts of the network. The main stated aim of the program to relieve key constraints in the network. Anecdotal evidence points to the view that the program was especially sensitive to constraints faced by the country's agricultural exporters (especially soy and corn) seeking to connect processing facilities to ports of export. Practically all investments were upgrades rather than new road construction. While some investments served to upgrade the road surface type (e.g. paving a dirt road) others improved the surface condition (e.g. repaving, filling pot holes, signaling) to improve the performance and capacity of existing roads.

An important concern is that the allocation of road upgrades to particular regions may be correlated with the expected performance of the local firms. This is a valid concern because

---

[23]Brazil ranks 123rd out of 144 countries on the World Economic Forum's "Quality of overall infrastructure" index, well behind China (51st) and India (74th). Surveyed executives cite an inadequate supply of infrastructure as the fourth most problematic factor for doing business, after tax rates, restrictive labor regulations, and corruption.

[24]Source: http://www.pac.gov.br/. The statistics refer to Phases I and II of the program which took place during 2007–2014. Phases III and IV are ongoing. Appendix Figure A.2 illustrates the extent and locations of the road investments programmed under PAC.

[25]For example China's National Trunk Highway Development Program whose stated objectives were to connect all major provincial capitals and cities (Faber 2014) or the Golden Quadrilateral Project which connects the four major cities in India (Goswami, Ghani and Kerr 2016).

spurring economic growth in certain regions is often a driver of infrastructure investment decisions. I address the endogenous road placement concern with the inclusions or a rich set of industry- and municipality- fixed effects in the empirical analysis. Specifically, all regressions include dummy variables for the roughly 250 industries and then more than 3,000 municipalities represented in the sample. These control for potential biases that would arise if the government targeted road investments toward particular industries throughout the country or specific municipalities that were expected to perform well (or poorly). Note that the inclusion of these fixed effects place a significant hurdle on the empirical analysis, as these fixed effects alone consume three quarters of the variation in the change in co-location measure.[26]

My identification strategy, then, exploits only the remaining variation, which is generated from *differences in pre-existing location patterns* across industries. The identifying assumption is that this within-industry-location variation is exogenous. The intuition behind the strategy can best be illustrated with a simple example (see Appendix Figure A.6). Consider two firms, one producing ice cream and the other soft drinks in the Brazilian municipality of Uberlandia. Assume that in 2007, other ice cream producers in the vicinity of Uberlandia happened to lie mostly to the West while other soft drinks manufacturers were mostly to the East. If Uberlandia saw an upgrade on a road leading westward, the ice cream producer would see a larger shock than the soft drink manufacturer. Thus, the identification strategy exploits industry-level differences in the pre-existing location patterns together with the municipality-level variation reflecting by the precise way in which road investments during 2007–2017 affected the travel time between the focal municipality and other municipalities in its four hour radius. As one test of the identification strategy, I check that the average change in co-location at the municipality-industry level calculated as above is uncorrelated with the firm survival trends in the preceding seven-year period, that is 2000–2007. Figure 1.3, a binscatter of the change in co-location against the pre-period survival rate, shows no

[26]Specifically, an OLS regression of change in co-location on industry and municipality fixed effects has an R^2 of 0.75.

significant relationship between firm survival rates in the pre-period and the subsequent change in co-location.

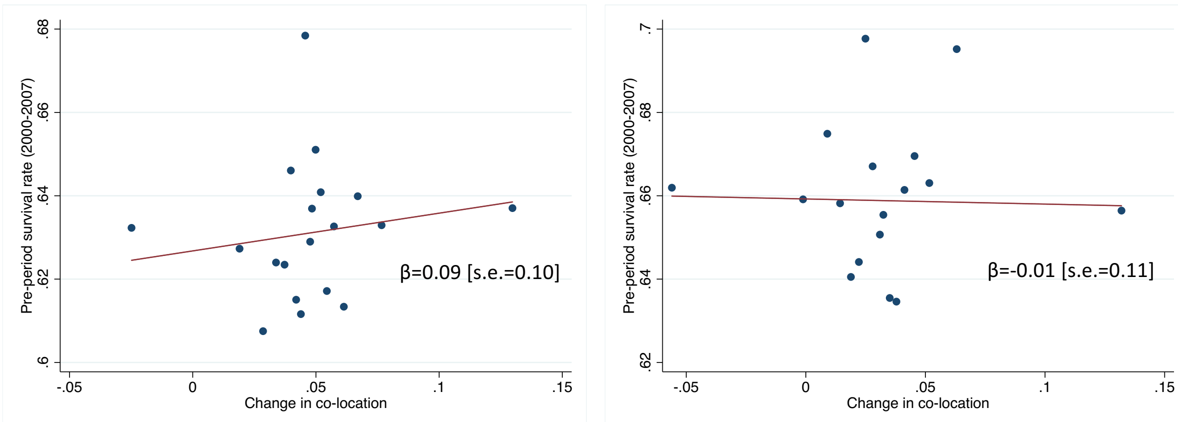

Figure 1.3: *Pre-period Survival Rates in Locally- (left) and Nationally Traded Industries (right)*
Note: Own calculations based on data from the Ministry of Transport of Brazil and RAIS.

1.5 Empirical Specification

To test the prediction of heterogeneous effects across the two industry types, I estimate the effects of increased co-location separately in the sample of locally traded and nationally traded industries. In each case, I estimate the following model:

$$Y_i = \beta_1 \Delta CoLoc^{sm} + \beta_p(\Delta CoLoc^{sm} \cdot Size07_i) + \beta_r x_{i,07} + \beta_q z^{sm} + ind + muni + \varepsilon_i \qquad (1.2)$$

Where Y_i is the outcome being studied, $\Delta\, CoLoc^{sm}$ is the change in co-location for industry s in municipality m calculated per Equation (1.1), x_i are firm-level controls in the baseline year, z^{sm} are industry-municipality level controls, all measured in the baseline year, ind and $muni$ are industry and municipality fixed effects, and ε_i is a randomly-distributed error term.

I estimate the model as a linear probability model (LPM) using ordinary least squares

(OLS) due to the large number of fixed effects.[27] In the simplest model, β_1 provides the estimate of the average effect of co-location on firm survival. However, in order to test the theoretical prediction of heterogeneous effects, I include interaction terms for the change in co-location and proxies for productivity that are based on firm size in 2007.[28] The first proxy is the simple measure of size (workers), $Size07_i$. As other proxies, I calculate different quantiles of firm size for each focal firm relative to all other firms in its industry, as well as relative only to firms in its industry and local market. In models that include the interactions of the co-location change with firm size, β_1 shows the effect for the omitted category (usually the smallest firms) while β_ps estimate the effects for the other firm categories.

1.5.1 Dependent Variable

The main dependent variable, Y_i, is firm survival over the seven-year period from 2007–2014. The advantage of this variable is that is most closely captures the theoretical mechanisms described in the models of spatial competition and firm selection whereby the least productive firms exit. Therefore, firm survival is the natural empirical outcome to be tested. I define the variable *Survive* as an indicator taking the value 1 if a 2007 incumbent firm continues to exist in the same municipality and four-digit industry in 2014 and a value of zero otherwise. Among the firms in the sample, the average seven-year survival rate is 55.6 percent — just over half of all firms. Arguably survival is a much cruder measure of the productivity growth theorized in the agglomeration literature where a more appropriate theory-backed measure would be residual TFP or value added.[29] I use firm survival as the

[27]A logit model, with more than 3,500 fixed effects, is unable to converge. For robustness, I have performed the analysis using a logit model, leaving out the municipality fixed effects (including in lieu state fixed effects). The direction and significance of the estimated coefficients is in line with the regressions performed using the LPM. Also, the estimated probabilities estimated with the LPM rarely fall outside the meaningful [0,1] interval.

[28]A strong positive relationship between firm size and productivity of manufacturing firms is well documented, for example in Haltiwanger, Lane and Spletzer (1999).

[29]Unfortunately, the data required to estimate TFP are not readily available for the cross-section of manufacturing firms in Brazil because such data are collected only for firms larger than 30 employees. In addition, TFP estimation has to confront numerous challenges, for example endogenous input choice. As a result, the agglomeration literature has in the past resorted to the use of employment growth as a productivity proxy, for example in Glaeser el al. (1992) and Henderson et al. (1995).

primary proxy of productivity growth, assuming that if firms are becoming increasingly productive they also are more likely to survive.

1.5.2 Controls

The baseline specifications include three types of controls: firm level, municipality level, and municipality-industry level controls. All are measured in the baseline year, 2007 and included in log-form. Table A.6 in the Appendix shows the correlations between the main variables.

Firm level controls: I control for firm size (workers) and firm age cohorts. Larger and older tend firms have higher survival rates and may have more power to attract preferential policies from the government. I also control for three other firm characteristics that are likely to be correlated with performance and which may make firms more likely targets of government policies. These are the firm's average worker wage (in reals), and dummies for whether the firm is an exporter or importer.

Municipality-industry level controls: All regressions control for the baseline level of a firm's co-location in 2007 for two reasons. One is that the baseline-level of co-location may be correlated with unobserved firm characteristics due to better firms sorting into more or less competitive locations or because more or less competitive locations result in more productive firms (e.g. though selection). Second, the baseline level of co-location is mechanically correlated with the change in co-location because of convergence effects (relative changes are smaller from a larger baseline). I also control for urbanization economies, measured as a count of firms in the focal firm's local market but outside its industry. Finally, I include a control for competition in the market-industry (measured as the inverse of the standard Herfindahl index) in order to control for any potential correlation between the change in co-location and trends in industry consolidation.

Industry fixed effects: All regressions include industry fixed effects in order to control for any potential correlations coming from macro-level industry shocks or the possibility that the government targeted road investments to specific industries throughout the country based on their expected future performance.

Municipality fixed effects: Finally, all regressions also include fixed effects for each of the more than 3,000 municipalities. These control for all unobserved characteristics that affect the performance of all manufacturing firms in the municipality during the 2007–2014 period equally (e.g. local shocks), and the possibility that the government targeted investments to certain well-preforming (or poorly-performing) regions.

1.6 Results

1.6.1 Co-Location and Survival in Locally Traded Industries

Table 1.1 shows the results of estimating Equation 1.2 in locally traded industries. The top rows show the main coefficients of interest, those on the change in co-location and its interaction with the productivity proxies (firm size and size quantiles). In interpreting the magnitude of the coefficients, $\beta * 100$ is the effect of a doubling in co-location (because change in co-location is a log difference, comparable to a percentage change). All non-categorical independent variables enter in log form, hence their coefficients can be interpreted as the estimated effect of a percentage point increase in the independent variable on the probability of survival, holding the other variables constant. All specifications include firm level controls in the baseline year, the industry-municipality, industry- and municipality fixed effects. Standard errors are clustered at the industry-municipality level to account for the fact that the co-location measure varies at that level.

The simple relationship between the change in co-location and firm survival is large and negative but not statistically significant in Column (1). Taken at face value the coefficient size suggests that a doubling in co-location leads to a 9.1 percentage point lower probability of survival, an elasticity of roughly 1/10. Column (2), in a first test of the prediction of heterogeneous firm-level effects, interacts the change in co-location with firm size. The coefficient on the change in co-location variable is now large, negative, and significant and the interaction effect is positive and highly significant. These results provide the first evidence that the effect of co-location is heterogeneous in firm size in locally traded industries.

Table 1.1: *Locally Traded Industries: Effects of Change in Co-Location on Firm Survival during 2007–2014*

Dependent variable: *Survive*	**Baseline** (1)	**Interacted with size** (2)	**2 Quantiles in industry** (3)	**3 Quantiles in industry** (4)	**4 Quantiles in industry** (5)	**4 Quantiles in ind-mkt** (6)
$\Delta\, CoLoc_{07-14}$	-0.091	-0.194***	-0.124**	-0.149**	-0.161**	-0.141**
	(0.059)	(0.068)	(0.062)	(0.065)	(0.069)	(0.071)
$\Delta\, CoLoc_{07-14}$ x $Size_{07}$		0.056***				
		(0.021)				
$\Delta\, CoLoc_{07-14}$ x 2nd Quantile			0.113**	0.119*	0.069	0.033
			(0.050)	(0.061)	(0.074)	(0.078)
$\Delta\, CoLoc_{07-14}$ x 3rd Quantile				0.115*	0.133*	0.053
				(0.066)	(0.076)	(0.084)
$\Delta\, CoLoc_{07-14}$ x 4th Quantile					0.159**	0.167**
					(0.076)	(0.079)
Firm level controls (2007)						
Number of workers (log)	0.059***	0.057***				
	(0.002)	(0.002)				
Firm in 2nd quantile			0.113**	0.082***	0.072***	0.073***
			(0.050)	(0.006)	(0.007)	(0.007)
Firm in 3rd quantile				0.136***	0.115***	0.119***
				(0.007)	(0.009)	(0.008)
Firm in 4th quantile					0.157***	0.152***
					(0.008)	(0.008)
Firm born prior to 1993	0.089***	0.090***	0.098***	0.095***	0.094***	0.095***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Firm born during 1993-2000	0.058***	0.059***	0.061***	0.059***	0.059***	0.060***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)
Average annual worker wage (log)	0.029***	0.029***	0.038***	0.034***	0.033***	0.033***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Exporter	-0.003	-0.003	0.023	0.019	0.017	0.018
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Importer	0.004	0.005	0.047***	0.042***	0.039***	0.041***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Municipality-Industry Controls (2007):						
Baseline co-location (log)	0.006	0.006	0.005	0.005	0.005	0.003
	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)
Urbanization economies (log)	0.137*	0.136*	0.151**	0.144**	0.146**	0.119
	(0.073)	(0.074)	(0.072)	(0.073)	(0.073)	(0.074)
Strength of competition (log)	0.003***	0.004***	-0.001	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Firms	52,496	52,496	52,496	52,496	52,496	52,496
Industries	73	73	73	73	73	73
Municipalities	2,982	2,982	2,982	2,982	2,982	2,982
R-squared	0.103	0.104	0.097	0.100	0.100	0.100
Industry FE	Y	Y	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y	Y	Y

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The table presents the long-differenced analysis of the linear probability model estimating the likelihood of firm survival during the seven year period 2007–2014 as a function of changes in co-location and baseline controls. Survive is an indicator variable taking the value 1 if a 2007 incumbent firm continues to exist in the same municipality and four-digit industry in 2014 and a value of zero otherwise. Identification comes from pre-period industry location patterns. Robust standard errors, clustered at the industry-municipality level, in parentheses. The models include the full set of industry and municipality fixed effects, thus some firms fall out of the sample. The omitted category in the case of size are firms in the first size quantile (smallest) and firms born after 2000 in the case of firm age.

Columns (3)–(5) report the results of the non-parametric model, interacting the change in co-location with progressively more granular quantiles of a firm's size in its industry. In each case the coefficient on the change in co-location, which measures the effect for the smallest firms (omitted category), is large, negative, and significant. Its size implies that a doubling of co-location reduces the survival probability of the smallest firms between 12 and 16 percentage points. Meanwhile, the coefficients for the larger quantiles are positive and significant, and roughly equal in size, suggesting no, or small, negative effects on survival for the biggest firms.

Column (6), where the quantiles are measured relative only to other firms in the focal firm's industry and market, provides evidence that the survival rates of the largest firms in the market increase as co-location increases. The marginal effects of this model suggest that doubling co-location, reduces the survival probability of the smallest firms by 14.1 percentage points and increases the survival rate of the largest firms by 2.6 percentage points.

Beyond the theorized effects, the coefficients of other variables in the model conform to expectations. Firm size and age wage have a positive and significant relationship with survival. The coefficient on firm size suggests that a doubling in firm size is associated with a roughly 6 percentage point higher survival rate. Older firms have higher survival rates than younger firms. Higher average wages are also predictive of higher survival while exporters and importers are, somewhat surprisingly, no more likely to survive in the baseline regression than non-trading firms. The effects of urbanization economies appear as positive and significant, suggesting more economic activity outside the own industry is associated with a higher survival rate, while the coefficients on the baseline levels of co-location and competition in the market and industry are small and not statistically significant.

Overall, the results point to *significant, negative effects of increased co-location on the smallest firms in locally traded industries*, lending support for Hypothesis 1. The results in column (6) also lend support for Hypothesis 2, showing *positive effects of increased co-location for the* *largest firms in locally traded industries*. Combined the results provide evidence of selection and reallocation among firms in the same local market and industry.

1.6.2 Co-Location and Survival in Nationally Traded Industries

Table 1.2 shows the parallel results for nationally traded industries. The coefficients on all the control variables are similar and therefore only the main results are presented. The differences in the main results are striking. In nationally traded industries, the baseline effect of a change in co-location on firm survival is positive, significant, and large. The coefficient in Column (1) suggests that in nationally traded industries, a doubling of co-location increases the survival rate by 14.9 percentage points, on average. Column (2), which interacts the shock with firm size, shows a negative but not statistically significant effect, providing weak evidence that the positive effect is smaller for larger firms. This conclusion is again confirmed in columns (3)-(6) where across the specifications, the results support the conclusion that doubling co-location increases the survival rate of the smallest firms by 18 to 20 percentage points. The coefficient on the interaction effects for the larger firms continue to be negative but, due to large standard errors, not statistically significant. The coefficients on the control variables in this sample are similar to those in the locally traded sample.

The baseline results provide significant evidence of *large, positive effects of increased co-location in nationally traded industries, with no significant differences across the firm size distribution*. Firms of all sizes benefit from increased proximity in locally traded industries. These results lend significant support for Hypothesis 3.

1.6.3 Relocations and Product Switches

In this section, I investigate whether firms react strategically to changes in co-location. A firm facing increased competition in its product and local market can respond by repositioning (Wang and Shaver 2013) for example by changing location or switching products. Given that they face the more serious competitive threats, I expect that the smallest firms in locally traded industries would be most likely to respond to increased co-location by repositioning, relocating to a different municipality or switching to a different product. Meanwhile, in nationally traded industries, increased co-location that increases spillovers makes it less likely that a firm would move closer toward others in its industry to seek spillovers or resources. Therefore, we should observe fewer moves following an increase in co-location due to the road shock.

Table 1.2: *Nationally Traded Industries: Effects of Change in Co-Location on Firm Survival, 2007–2014*

Dependent variable: *Survive*	Baseline (1)	Interacted with size (2)	2 Quantiles in industry (3)	3 Quantiles in industry (4)	4 Quantiles in industry (5)	4 Quantiles in ind.-mkt. (6)
$\triangle\ CoLoc_{07-14}$	0.149**	0.217**	0.201**	0.188**	0.178*	0.184*
	(0.076)	(0.110)	(0.083)	(0.092)	(0.103)	(0.106)
$\triangle\ CoLoc_{07-14}$ x $Size_{07}$		-0.030				
		(0.038)				
$\triangle\ CoLoc_{07-14}$ x 2nd Quantile			-0.106	0.035	0.063	-0.063
			(0.097)	(0.118)	(0.112)	(0.116)
$\triangle\ CoLoc_{07-14}$ x 3rd Quantile				-0.121	-0.030	-0.023
				(0.113)	(0.134)	(0.134)
$\triangle\ CoLoc_{07-14}$ x 4th Quantile					-0.133	-0.122
					(0.134)	(0.145)
Firm level controls (2007)						
Number of workers (log)	0.054***	0.055***				
	(0.004)	(0.004)				
Firm in 2nd quantile			0.108***	0.107***	0.089***	0.078***
			(0.009)	(0.010)	(0.011)	(0.011)
Firm in 3rd quantile				0.152***	0.147***	0.139***
				(0.011)	(0.011)	(0.011)
Firm in 4th quantile					0.169***	0.153***
					(0.013)	(0.013)
Firm born prior to 1993	0.083***	0.083***	0.089***	0.086***	0.086***	0.087***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Firm born during 1993-2000	0.043***	0.043***	0.046***	0.044***	0.044***	0.045***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Average annual worker wage (log)	0.022***	0.022***	0.030***	0.026***	0.025***	0.026***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Exporter	0.001	0.001	0.026*	0.022	0.022	0.022
	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)
Importer	0.019	0.020	0.047***	0.043***	0.044***	0.046***
	(0.015)	(0.015)	(0.014)	(0.014)	(0.015)	(0.015)
Municipality-Industry Controls (2007):						
Baseline co-location (log)	0.014**	0.014**	0.016**	0.014**	0.014**	0.015**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Urbanization economies (log)	0.092*	0.091*	0.094*	0.093*	0.090	0.085
	(0.055)	(0.055)	(0.057)	(0.056)	(0.056)	(0.055)
Strength of competition (log)	0.001	0.001	-0.003*	-0.001	-0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Firms	21,930	21,930	21,930	21,930	21,930	21,930
Industries	99	99	99	99	99	99
Municipalities	2,302	2,302	2,302	2,302	2,302	2,302
R-squared	0.103	0.104	0.097	0.100	0.100	0.100
Industry FE	Y	Y	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y	Y	Y

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. The table presents the long-differenced analysis of the linear probability model estimating the likelihood of firm survival during the seven year period 2007–2014 as a function of changes in co-location and baseline controls. Survive is an indicator variable taking the value 1 if a 2007 incumbent firm continues to exist in the same municipality and four-digit industry in 2014 and a value of zero otherwise. Identification comes from pre-period industry location patterns. Robust standard errors, clustered at the industry-municipality level, in parentheses. The models include the full set of industry and municipality fixed effects, thus some firms fall out of the sample. The omitted category in the case of size are firms in the first size quantile (smallest) and firms born after 2000 in the case of firm age.

In further tests, I analyze product switching and firm relocation. I define the variable *Switch* as an indicator taking the value 1 if an incumbent firm reports a different industry code in 2007 and in the last year that it is observed in the sample. Similarly, I define the dummy variable *Relocate* if an incumbent firm is located in a different municipality the last year that it is observed.[30]

I explore these predictions by analyzing whether the prevalence of product switches and firm relocations was affected by changes in co-location stemming from the road upgrades. The model is parallel to Equation (1.1) but, in addition, the firm level controls now also include a control for the last year that the firm is observed in the sample, i.e. the year of "exit" (or 2014 in the absence of exit). This control is important because time is the main predictor of product switches and relocations and, as we know from the prior analysis, changes in co-location affect the likelihood of firms surviving and thus remaining in the sample.

Table 1.3 shows the results of the analyses of moves and relocations in locally and nationally traded industries. The differences across the two industry types are once again impressive and in line with theory. In locally traded industries, the results for the sample as a whole show that increases in co-location increase the likelihood of relocating and switching to a different industry, although the effects are not significant at typical significance levels (Columns 1 and 3). The results become clearer after introducing the interactions with firm size quartiles.

For the smallest firms, the likelihood of *moving or switching industry increases* as co-location increases. Specifically, doubling co-location increases the likelihood of moving by 3.8 percentage points and the likelihood of industry switching by 7.4 percentage points. Given the average likelihood of moves and product switches of 3.1 and 8.4 percent, respectively, this is roughly a doubling of the likelihood of these events. Again, larger firms are less

[30] Note that, given the definition of *Survive*, any firm moving or switching is part of the subset of firms defined as having not survived (in their original location and industry). The main results on survival are robust to the dropping of movers and switchers from the sample, i.e. full "exits".

likely to move and relocate, through the differences are not always statistically significant.

In nationally traded industries, on the other hand, the likelihood of *moving and switching industry falls* with an increase in co-location. For the smallest firms, doubling co-location decreases the probability of moving by 5.2 percentage points, again, more than doubling the baseline probability of 4.1 percent. The coefficients on the interaction terms suggest that the effect is muted for large firms, though large standard errors render the interaction not statistically significant. The result on switches, which were not clear in theory, is also more ambiguous in the empirical results. While all coefficients are negative, suggesting lower propensity to switch industries, none are statistically significant at the standard thresholds.

Table 1.3: *Effects of Change in Co-Location on Relocations and Product Switches*

	Locally traded				Nationally traded			
Dependent variable:	*Move*		*Switch*		*Move*		*Switch*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta\, CoLoc_{07-14}$	0.027	0.038*	0.056	0.074*	-0.042	-0.074**	-0.065	-0.064
	(0.019)	(0.022)	(0.039)	(0.041)	(0.031)	(0.031)	(0.050)	(0.061)
$\Delta\, CoLoc_{07-14}$ x 2nd Quantile		-0.037**		-0.053		0.040		-0.002
		(0.017)		(0.036)		(0.041)		(0.078)
$\Delta\, CoLoc_{07-14}$ x 3rd Quantile		-0.001		-0.021		0.073*		0.006
		(0.021)		(0.039)		(0.044)		(0.070)
Firm in 2nd quantile		0.009***		0.027***		0.008**		0.027***
		(0.002)		(0.004)		(0.004)		(0.006)
Firm in 3rd quantile		0.015***		0.039***		0.018***		0.035***
		(0.002)		(0.004)		(0.005)		(0.008)
Firms	52,496	52,496	52,496	52,496	21,930	21,930	21,930	21,930
R-squared	0.077	0.078	0.113	0.115	0.114	0.115	0.152	0.154
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y
Municipality FE	Y	Y	Y	Y	Y	Y	Y	Y

*** p<0.01, ** p<0.05, * p<0.1. The table presents the long-differenced analysis of the linear probability model estimating the likelihood of relocations and product switches during the 2007-2014 period as a function of the changes in co-location and baseline controls. Move/Switch as an indicator taking the value 1 if a 2007 incumbent firm is observed in a different municipality/ industry code in the last year that it appears in the data. Identification comes from pre-period industry location patterns. Robust standard errors, clustered at the industry-municipality level, in parentheses. The omitted category in the case of size are firms in the 1st size quantile (smallest) and firms born after 2000 in the case of firm age. All specifications include the full set of controls.

A final difference becomes apparent considering the destinations that firms move to when relocating or switching industries. I calculate, for all firms that move and switch industry, the difference in co-location between the origin and destination. In Figure 1.4, we see that firms that relocate or switch industries in locally traded industries tend to move *away* from competitors, while in nationally traded industries, they move *toward* competitors. While this last piece of evidence is descriptive, it lends support for the main hypothesis of the paper, that proximity to competitors plays a fundamentally different role in nationally traded and locally traded industries.

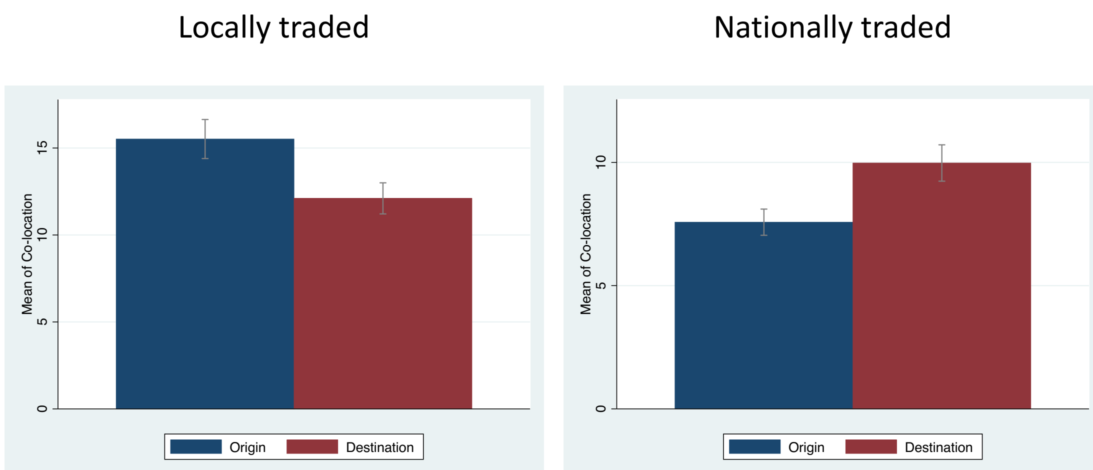

Figure 1.4: *Origin and Destination Co-Location for Firms that Move and Switch*
Note: Own calculations based on data from the Ministry of Transport of Brazil and RAIS.

1.6.4 Robustness

A key challenge lies in distinguishing the effects of increased co-location with competitors from other potential effects of improved roads, e.g. improved access to customers, changes in the prices of inputs, or increased entry of new firms. In this section I test the robustness of the main findings to these possibilities.

Establishing the effect that improved roads may have on firms through the demand channel — i.e. by enabling better access to markets and customers — has been the focus of number of existing studies (e.g. Donaldson and Hornbeck 2016). The standard approach to capturing changes in market access is as the GDP or population accessible to a firm. I create a variable named change in local market access, which is the change in travel-time weighted sum of the municipal GDP off all municipalities within a firm's four-hour market from 2007 to 2014. As in the measure of competitor access, variation comes from travel time changes and changes in the market area (not the GDP, which is kept at its 2007 value). Note that unlike the competitor access measure, market access varies at the municipality level and hence can only be included in a model that excludes municipality fixed effects.

Table 2.7 shows the results, comparing the baseline model to a model that excludes the municipality fixed effects (Column 2) and then adds a control for the change in local market access (Column 3). For the locally traded industries, both the removal of the municipality fixed effects and the inclusion of the additional control do little to change the main result of a strong negative effect on the smallest firms. The only difference is that the coefficient for the largest firms is now smaller in magnitude and no longer statistically significant. In nationally traded industries, however, the parallel models in Columns (6) and (7) are sensitive to the removal of the municipality fixed effects but do not appear sensitive to the inclusion of the market access control. The sensitivity to the fixed effects suggest that road investments were targeted to regions with *worse* prospects in traded sectors, confirming the importance of the municipality controls.

Columns (4) and (9) check whether there is evidence that the results are driven by the entry of new firms. It may be that road improvements attracted new entrants and that both the competitive and agglomeration effects seen by the incumbent firms are a result of the change in entry dynamics. Controlling for the 2007–2014 entry rate, calculated as the number of new entrants during 2007–2014 as a share of the 2007 incumbent firms, the results provide evidence that higher entry has a statistically significant negative association with survival probability for incumbents, suggestive of increased competition. However, the inclusion of the entry control does not affect the main effect of the change in co-location.

Finally, Columns (5) and (10) look for evidence of increased factor prices, by replacing firm survival with a new dependent variable which is the growth in the average wage of the firm from 2007 to 2014. Note that this analysis can only be performed on the set of

Table 1.4: *Robustness Checks*

	Locally traded industries					Nationally traded industries				
	Baseline	Without Muni FE	Local mkt. control	Entry control	Factor prices	Baseline	Without Muni FE	Local mkt. control	Entry control	Factor prices
Dependent variable:	*Survive*	*Survive*	*Survive*	*Survive*	*Wage growth*	*Survive*	*Survive*	*Survive*	*Survive*	*Wage growth*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\Delta\, CoLoc_{07-14}$	-0.161**	-0.115**	-0.141**	-0.168**	0.028	0.178*	0.051	0.060	0.175*	-0.163
	(0.069)	(0.054)	(0.061)	(0.070)	(0.381)	(0.103)	(0.084)	(0.087)	(0.103)	(0.468)
Δ *Local mkt. access*			0.026					-0.014		
			(0.030)					(0.045)		
$Entry\ rate_{07-14}$				-0.018***					-0.011*	
				(0.003)					(0.006)	
$\Delta\, CoLoc_{07-14} \cdot Quant2$	0.069	0.036	0.034	0.069	-0.424	0.063	0.020	0.020	0.062	-1.018
	(0.074)	(0.070)	(0.071)	(0.074)	(0.469)	(0.112)	(0.110)	(0.110)	(0.112)	(0.730)
$\Delta\, CoLoc_{07-14} \cdot Quant3$	0.133*	0.092	0.089	0.135*	-0.174	-0.030	-0.091	-0.091	-0.029	0.014
	(0.076)	(0.077)	(0.077)	(0.076)	(0.414)	(0.134)	(0.116)	(0.117)	(0.134)	(0.505)
$\Delta\, CoLoc_{07-14} \cdot Quant4$	0.159**	0.114	0.111	0.160**	-0.356	-0.133	-0.129	-0.129	-0.131	0.008
	(0.076)	(0.072)	(0.073)	(0.076)	(0.487)	(0.134)	(0.121)	(0.121)	(0.134)	(0.531)
Firms	52,496	52,985	52,985	52,496	28,870	21,930	22,575	22,575	21,930	11,849
R-squared	0.100	0.054	0.054	0.101	0.128	0.152	0.071	0.071	0.152	0.186
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Municipality FE	Y	N	N	Y	Y	Y	N	N	Y	Y
State FE	N	Y	Y	N	N	N	Y	Y	N	N

surviving firms. The results do not show evidence that the shock to co-location affected wage trends in incumbent firms, offering little evidence for the input prices channel in the case of labor.

1.7 Conclusion

This paper estimates the effect of co-locating with firms in one's industry on firm survival by leveraging reductions in travel times between firm locations stemming from improved roads as exogenous variation in co-location. I find that in industries that compete for customers locally, increased co-location produces effects consistent with heightened competition: doubling co-location lowers the survival rate of the smallest firms by 14.1 percentage points while increasing it by 2.6 percentage points for the largest firms. In industries that compete in national markets, increased co-location produces effects consistent with increased agglomeration spillovers. Doubling co-location increases firms' survival rate by 14.9 percentage points, with no significant differences across firm sizes. As further evidence, consistent with increased competition in locally traded industries, I observe a higher propensity of firms to move to a different municipality or switch their primary product after being brought closer to competitors, and when they do so, they tend to evade competition. Meanwhile, in nationally traded industries I observe fewer relocations and when these occur, they are moves towards competitors.

The findings of significant differences in the way that firms respond to co-location with competitors suggest that there is not a single answer to the question how proximity affects performance, but rather that studies need to be careful to consider both industry and firm level heterogeneity. While the focus of this study is in establishing that first-order dimension of heterogeneity, one limitation that stems from its cross-industry nature is the inability to point to the relevance of specific mechanisms behind the effects (e.g. spillovers from shared input suppliers versus richer labor pools). These could be evaluated in future work with a narrower scope, e.g. studies in a single industry or region.

The ability of the study to detect effects on firm behavior based only on changes to the actual cost of mobility also carries the important implication that "space" is not a constant but rather is shaped by the costs and patterns of human mobility. This opens up opportunities for further inquiry on how changes in technologies and policies that affect the cost of mobility shape the competitive and collaborative interactions between firms.

Chapter 2

The Internal Agglomeration of Multibusiness Firms

2.1 Introduction

[1]Multibusiness firms account for an oversized share of economic activity and, given their practical importance, have inspired a large body of research in economics, finance, and strategy. The resource-based view sees the ability of firms to generate synergies by sharing resources simultaneously among business units (Penrose 1959, Teece 1982) and their ability to redeploy resources among business units over time in response to changing conditions (Helfat and Eisenhardt 2004, Levinthal and Wu 2010) as a key driver of the diversification decision itself and an important source of excess value and competitive advantage of multibusiness relative to stand-alone firms (Sakhartov and Folta 2014).

While such competitive advantages hinge on resources actually being exchanged among the different units of the firm, the corporate strategy literature is largely silent on the role that the *spatial organization* of the firm plays in enabling or constraining resource flows. At the same time, several recent studies have found that geographic distance between a firm's units affects the effective allocation of capital (Giroud 2012, Giroud and Mueller 2014, Choudhury 2017) and managerial attention (Mingo 2013), as well as the extent of input-output flows (Atalay, Hortacsu, Li and Syverson 2017) and knowledge flows (Lahiri 2010, Keller and Yeaple 2013). The finding that geographic proximity is a key enabler of

[1]This chapter is joint work with Juan Alcácer, Harvard Business School.

resource flows within firms begs the question: do multibusiness firms locate in ways that facilitate synergies and resource redeployment among their different business units?

In this paper, we examine the spatial organization of complex, multibusiness firms and ask to what extent the potential for resource flows among their business units affects their geographic proximity inside the firm. Leveraging the insight that geographic proximity is an enabler of resource flows, we propose and test the hypothesis that multibusiness firms exhibit "internal agglomeration" — a systematic and predictable pattern of co-location of business units *within* the firm (Alcácer and Delgado 2016). We distinguish internal agglomeration from the well-known phenomenon of industry agglomeration (which henceforth, we will term "external" agglomeration), which refers to co-location *between* unrelated firms. The literature studying the tendency of industries to agglomerate externally[2] has found that in many industries, firms co-locate more than would be expected at random given population location patterns (Ellison and Glaeser 1997, Duranton and Overman 2005).

Most existing agglomeration studies abstract from differences in firms' organizational form – in particular, whether a firm consists of multiple business units over which it is optimizing its location decision or whether it's a stand-alone, single-location firm. In this paper, we explicitly separate the two types of firms. While we focus our analysis on multibusiness firms, we use the external agglomeration of stand-alone firms in the same industries as a benchmark against which to compare the tendency of firms to agglomerate *internally*. We hypothesize that, compared to the agglomeration of stand-alone firms, the tendency to agglomerate internally will be higher, because internal proximity enables resource sharing without generating negative effects at the firm-level from the increased competition over local resources and customers (Baum and Haveman 1997).

However, if internal agglomeration benefits were the only consideration in a firm's location choice, we would observe that firms concentrate all activities in one location. This is, of course, not the case, since firms have other objectives that encourage them to expand

[2]In the literature, the term "agglomerate" refers to the tendency of firms in the *same* industry to co-locate while the term "coagglomerate" refers to the tendency of firms in *different* industries to co-locate. In this paper, we use the term "agglomerate" generally, to refer to both within-industry and across-industry co-location.

geographically, for example, accessing new markets and more distant resources. Unless the benefits from co-locating are large, these alternative objectives will push the boundary of the firm outward. Corporate strategy literature argues that firms are more likely to benefit from synergies and the flexibility of resource redeployment when they are active in industries that are more resource-related (Rumelt 1982, Montgomery and Hariharan 1991, Farjoun 1994, Helfat and Eisenhardt 2004). Because relatedness is a necessary condition for resource sharing to be desirable, we propose that internal agglomeration will be highest among business units in industries that are resource-related.

Assuming that two industries are resource-related, another important question is how sensitive the particular type of resource being exchanged is to geographic distance. It's important to note that resource relatedness between industries can exist along multiple dimensions. For example, some industries are similar in their labor requirements. Others employ different types of workers but use similar technologies. If workers are relatively immobile compared to technology, for example, then industries that are related in the labor dimension are more likely to co-locate than industries related in technology. This insight, which we apply to uncover the drivers of internal agglomeration in multibusiness firms, has also been leveraged to study the importance of different drivers of external agglomeration (Ellison, Glaeser and Kerr 2010).

We hypothesize that labor relatedness will have the strongest effect on internal agglomeration. Workers are a non-scale free resource and if a firm wants to leverage a worker in multiple business units simultaneously, geographic proximity is key. If workers are to be redeployed, moving over greater distances is much costlier than reallocating a worker to a nearby business unit. Meanwhile, we hypothesize that the desire to ship intermediate inputs between business units is a less important predictor of internal agglomeration. First, recent evidence calls into question to what extent vertically-related business units owned by the same firm actually buy and sell from one another (Atalay, Hortaçsu and Syverson 2014). Second, the need to monitor a supplier is lower when the supplier is integrated compared to when the parties are unrelated. Hence, we expect input-output linkages to be less predictive of internal agglomeration relative to their role in external agglomeration. Finally, we believe that relatedness in technology or knowledge will exert smaller effects on internal relative to external agglomeration. Although some knowledge resources are likely to be quite localized (e.g. tacit knowledge embedded in local routines), others (e.g. patents) are likely to be public knowledge inside the firm. These are scale free resources and can be used at little costs in distant plants (Levinthal and Wu 2010). In addition, based on existing evidence, knowledge-sharing while sensitive to distance even within the firm, may be less so than between unrelated firms (Almeida, Song and Grant 2002).

We study internal agglomeration using establishment-level data from Dun & Bradstreet which provide the plant location (ZIP code), up to six different industry codes produced at the establishment, and complete corporate familiy tree linking each establishment to all others owned by the same parent for large sample of U.S. firms. We calculate our main measure of internal agglomeration with a modified Duranton-Overman index (Duranton and Overman 2005). Specifically, we measure distances among all plants belonging to the *same* firm. The measured distances between all plants of a firm provide the input with which we construct a measure of continuous agglomeration for each pair of different manufacturing industries. For each industry pair ij and distance threshold d, the internal agglomeration measure answers, what is the likelihood that two establishments of a multibusiness firm picked at random, one in industry i and the other in j are located d kilometers apart? Separately, we also measure distance among unrelated plants, i.e. the distances among single-product, single-location ("stand-alone") firms and with this, construct a measure of *external* agglomeration. External agglomeration serves as the benchmark against which we assess the degree of internal agglomeration in multibusiness firms. While the continuous index has many advantages and is our primary measure of agglomeration, for robustness, we also construct measures of agglomeration in discrete geographic units, i.e. the likelihood that two plants picked at random are located in the same county, the same core-based statistical area (CBSA), and the same state.

To test the mechanisms driving internal agglomeration we employ several measures of industry relatedness. One is the tendency of industries to buy and sell from one another (input-output linkages). The second is a measure of the similarity of industries' occupational requirements, which proxies for the potential of plants in different industries to share managerial and human resources (labor similarity). The third is a measure of the potential of two plants in different industries to benefit from the flow of knowledge, proxied by the extent to which an industry uses technology generated by the other (technology linkages).

Our results show that internal agglomeration is a general feature of the spatial organization of multibusiness firms. Across all industry pairs, the average tendency to locate two plants at close proximity inside a firm is about twice as high as the tendency of two unrelated firms to co-locate (which we also know is higher than random). For example, the expected probability that two randomly picked plants are at a distance of 250 kilometers or less is 10.2 percent if those two plants belong to the same multibusiness firm and 5.7 percent if they are plants of stand-alone firms.

Examining the drivers of agglomeration, we find that they are somewhat distinct in multibusiness and stand-alone firms. For multibusiness firms, among the three types of industry relatedness that we study, labor similarity exerts the largest and most significant effects on internal agglomeration. The effects of labor similarity are significant across all distance thresholds. Input-output linkages are statistically significant only beyond the 250-kilometer threshold, predicting that while input-output related plants don't necessarily agglomerate at short distances, they tend to locate not too far apart. Third, the evidence points to the conclusion that technological linkages among industries are not a significant driver of internal agglomeration. This finding is consistent with the view that knowledge can be transferred at larger distances inside firm boundaries, thus lowering the incentive to co-locate in order to take advantage of knowledge spillovers.

Among the stand-alone firms, on the other hand, we find that all three measures of industry relatedness have positive and significant effects on agglomeration, in line with prior findings (e.g. Ellison, Glaeser and Kerr 2010). Considering the differences in internal versus external agglomeration levels explicitly, we find that labor similarity is the single most important factor driving the *excess* agglomeration of multibusiness firms. This finding suggests that internal labor markets are a potentially important mechanism inside multibusiness firms though which synergies and benefits from resource redeployment are realized, which is consistent with a recent literature that finds that multibusiness firms actively use (Tate and Yang 2015) and benefit from the flexibility offered by internal labor markets (Belenzon and Tsolmon 2016).

The main contribution of this study is to provide one of the first systematic analyses of the empirical patterns of the overall spatial organization of multi-location, multibusiness firms. We document the phenomenon of internal agglomeration across a wide set of firms and industries. While prior studies have considered certain aspects of the firm location choice, for example location of the first unit (Kalnins and Chung 2004), the location of foreign entrants (Shaver and Flyer 2000, Chung and Alcácer 2002) or headquarters (Bel and Fageda 2007, Strauss-Kahn and Vives 2009) this is, to our knowledge, the first large-scale analysis of the overall spatial relationship between the plants of multibusiness firms and as well as the drivers of internal agglomeration.

Our findings contribute to existing research in the geography and location strategy literature. This literature has studied the costs and benefits of agglomerating and clustering (Delgado, Porter and Stern 2014, Ellison, Glaeser and Kerr 2010), as well as heterogeneity in the benefits of agglomeration across firms (Shaver and Flyer 2000, Alfaro and Chen 2014), though it has almost exclusively focused on external agglomeration rather than agglomeration within the firm itself. Alcácer and Delgado (2016), who consider both internal and external agglomeration forces and first identify the "internal agglomeration" concept, is most related to our current study. In their study of the biopharmaceutical industry, the authors find that firms consider the location of their own plants as well as unrelated plants in the same industry in subsequent location choices. The current study generalizes this finding, documenting the phenomenon of internal agglomeration across many firms and industries, and sheds light on the mechanisms driving it.

We also contribute to the corporate strategy literature, especially the recent literature on resource redeployment (reviewed in Folta, Helfat and Karim (2016)). Our findings suggest that a firm's geographic footprint can enable but also constrain the extent to which a firm can engage in internal resource sharing and redeployment. In particular, our study points to the importance of geographic proximity as an enabler of human-resource driven economies of scope. Our findings suggest that there is an important interdependence between corporate strategy and location strategy of a firm and that an explicit consideration of the spatial dimension of the firm is important in studies of resource synergies and redeployment.

Our results also have implications for practitioners who manage complex multibusiness firms that frequently span multiple geographies. While the lowering of trade and transport barriers over the past few decades have pushed the geographic boundaries of the firm outward, face-to-face interactions and geographic proximity remain important. Understanding how to optimally manage the competing objectives of geographic expansion and internal proximity remains relevant.

2.2 Theoretical Background and Hypotheses

In this section, we review the key tenets of the corporate strategy literature on synergies and resource redeployment in multibusiness firms. We integrate these theories and recent findings from a variety of contexts that point to an important role of geographic proximity in intra-firm resource flows. From these two literatures we build our hypotheses, predicting to what extent different resource sharing objectives are expected to shape the location strategy in multibusiness firms.

2.2.1 Resource Flows in Multibusiness Firms

Multibusiness firms play an oversized role in the modern business landscape. In the U.S., they account for just under 30 percent of all firms but more than 80 percent of manufacturing output (Bernard, Redding and Schott 2010) and more than 90 percent of the value of exports (Bernard, Jensen, Redding and Schott 2007). Among publically listed firms, they account for about half of all assets, revenues, and market value (Folta, Helfat and Karim 2016). Their plants are larger and more productive than comparable plants of stand-alone firms (Schoar 2002). Workers inside multibusiness firms produce more output per unit of labor and earn higher wages than workers in stand-alone firms (Tate and Yang 2015).

Given their practical importance and apparent competitive advantages, the question why firms diversify into multiple businesses and what the sources of the competitive advantages of multibusiness firms are has given rise to a rich stream of research. The resource-based view sees the firm as "those (tangible and intangible) assets which are tied semipermanently to the firm" (Wernerfelt 1984, p. 172) and posits that multibusiness firms arise as vehicles for the mobilization of excess or "slack" resources that are fungible across activities but would otherwise be idle because they are difficult to transact in the open market (Penrose 1959, Teece 1982). These resources can be tangible (e.g. physical and financial capital) or intangible (e.g. tacit knowledge of workers, managerial experience, organizational routines). Assuming that the firm cannot infinitely expand in its primary product, it can diversify into other products and leverage its existing resources, generating synergies. The resources can best be leveraged in activities that are related, or similar to the firm's core business because these activities balance economies of scope with the diseconomies of organizational scale (Rumelt 1982).

Newer additions to this literature highlight that while some internal resources have the quality of being public goods within the firm (scale free) others are rival and their use in one part of the business limits their use in other parts (non-scale free) (Levinthal and Wu 2010). For example, a firm's brand is a scale free resource and can be simultaneously leveraged in different units of the firm. The time and skill of the firm's top managers is a non-scale free resource. Firms can achieve advantages in resource use even for resources that are non-scale free through resource redeployment (Helfat and Eisenhardt 2004). Rather than using resources simultaneously in two different business units (generating synergies, or intra-temporal economies of scope), firms can redeploy resources from activities where there are less valuable and toward activities where they are more valuable (generating inter-temporal economies of scope). Resource redeployment can help firms escape declining prospects or negative shocks in some industries in their portfolio while lowering the costs of starting-up (Lieberman, Lee and Folta 2016) or expanding operations in more promising areas. As is the case for synergies, the relatedness of the industries is a key determinant of the likelihood that resources will be in fact be redeployed. Notably, in either scenario, a *physical movement* of resources (e.g. equipment, workers, goods, ideas) is expected to take place between the different business units.

2.2.2 Geography and Intra-Firm Resource Flows

Despite the fundamentally geographic nature of many types of resource flows, questions related to the physical location of the multibusiness firms' units are not directly addressed by the corporate strategy literature. This is a potentially important omission. A number of recent papers provide direct empirical evidence that geographic proximity is a critical enabler of resource allocation within firms. For example, Schoar (2002) finds that after a firm diversifies into a new industry though an acquisition, the incumbent plants incur a decrease in productivity, while the acquired plant increases its productivity, suggesting that resources such as managerial attention are being transferred from existing to the acquired plant. Mingo (2013) finds a similar pattern in the sugar processing industry in Brazil and shows moreover that the decrease in the performance happens at those units that are geographically proximate to the acquired unit. Giroud (2012) documents that the allocation of capital from headquarters to subsidiaries increases in response to decreases in effective distance between the headquarter and the plant following the introduction of new airline routes. Giroud and Mueller (2014) further show that this reallocation toward the focal plant occurs along with a drawing of resources away from plants that are underperforming and more distant from the headquarters. Choudhury (2017) finds that the proximity of a business unit manager to the corporate headquarters impacts the allocation of resources toward innovative activities of that business unit.

The mounting evidence showing a direct causal relationship between geographic proximity and intra-firm resource flows begs the question to what extent enabling resource flows among the firm's business units shapes the overall location strategy and spatial organization of multibusiness firms. This is a topic that, thus far, has not been addressed either in the strategy or the economic geography literatures. In a review of the location strategy literature, McCann and Folta (2008) note that "the extant literature envisions firms as unitary actors with one location, when in reality firms often have multiple establishments in varied locations" (p. 552). Similarly, in a recent review of the agglomeration literature in economic geography, Kerr and Duranton (2015) write that "firm-level studies of how multi-unit firms interact with local agglomeration economies versus internally sourced resources are woefully few in number" (p. 3). In the next section of this paper we explicitly consider the unique spatial organization problem of multibusiness firms and how it differs from that of stand-alone firms.

2.2.3 Location Strategy in Multibusiness Firms

We start our discussion by contrasting the location decision of a multibusiness plant with that of a stand-alone plant. In choosing a location for its operations, a stand-alone plant faces a well-defined problem: where to locate its one (and only) plant. Prior studies have evaluated this problem, and theorized about the importance of different features of the *external* environment which may attract firms to a location, for example the local demand conditions, the availability of factors and inputs, the presence of other firms in the focal and related industries, the availability of local knowledge pools, for example universities and R&D labs, the quality of the local institutional environment (Porter 1990). Meanwhile the location problem of a multibusiness, multi-plant firm is more complex because of an additional consideration: the locations of its pre-existing plants. For a multibusiness firm, with the exception of the first plant, subsequent location choices will take into account not only the quality of the *external* environment but also *internal* considerations.

There are several theoretical reasons why multibusiness, multi-plant firms may wish to stay spatially compact. One set of reasons relate to the organizational costs of monitoring and managing geographically distant plants. In general, distance creates higher costs associated with communicating and exchanging information. Managers may need to make frequent trips to the various firm locations, incurring the costs of travel and the opportunity costs of time. Control systems may be weaker in geographically distant plants and monitoring more difficult (Kalnins and Lafontaine 2013). A firm may face the "liability of foreignness" as it enters new institutional environments (Zaheer 1995). All of these are sources of incremental costs associated with managing geographically distant firm units.

Beyond avoiding costs, however, geographic proximity of the firm's plants may also generate certain benefits. In particular, we posit that geographic proximity is a critical enabler of intra-temporal and inter-temporal economies of scope. Consider the example of a firm which has slack in the use of a certain machine, discussed in Teece (1982). Such slack is difficult to transact in the market, and will provide an incentive for the firm to diversify its product portfolio in order to leverage the resource slack and generate synergies. It is obvious in this example that the geographic location of this second activity is fully determined — given the excessive costs associated with moving a piece of machinery back and forth between activities, the second activity will necessarily take place *in the same plant*. Thus, the desire to achieve synergies in the use of a capital resource has strong implications for the geographic proximity of activities inside a firm in this example.

Consider on the other hand, a firm which generates synergies in is brand, by leveraging it across multiple product portfolios. A firm's brand is a non-scale free resource — it can be used simultaneously at multiple locations inside the firm. Moreover, using the brand does not pose any "transportation costs." Therefore, for this resource, the physical proximity of plants inside the firm is not required in order for the firm to benefit from economies of scope in the use of the brand.

These examples illustrate a key insight regarding the relationship between potential for synergies (intra-temporal or inter-temporal) in resource use, the "transportation cost" involved in using a resource at multiple plants and the implications thereof for the proximity of plants within the firm. The same line of thinking can be applied to many other examples of resources, for example human capital or patents. Overall, both the desire to avoid the costs of managing over large distances and the desire to reap the benefits of shared resource use may motivate the firm to co-locate its business units.

However, while there are these potential benefits, co-locating the firm's plants also has some opportunity costs. On the demand side, the desire to expand into new markets and locate close to final consumers motivates firms to expand horizontally. On the supply side, a firm may also wish to diversify its location in order to access certain inputs (e.g. natural resources). Finally, as for stand-alone firms, external agglomeration economies can also present a draw for the multibusiness firms to locate close to other unrelated firms.

Ultimately, multibusiness firms will have to trade off the potential benefits from co-locating internally against the opportunity costs of doing so. The "centripetal" force of lower organizational costs and easier resource sharing will drive the firm's units together, while the "centrifugal" forces of demand-, resource- and external agglomerations will drive the boundaries of the firm apart (Alcácer and Delgado 2016). How can we predict which force will dominate and whether the units within a firm will co-locate or not? Since the benefits of co-locating are a function of the extent to which industries can engage in resource sharing, the likelihood of co-location of any two business units should increase in the resource relatedness between them. The potential for units to generate synergies and to benefit from resource redeployment can be proxied by how resource related they are (Sakhartov and Folta 2014). The more related are two business units in the kind of resources that they use in production, the more likely they will be to generate synergies, and thus the more likely they are to be more geographically proximate inside the firm. This leads us to purpose that:

Hypothesis 1: Multibusiness firm will tend to co-locate (internally agglomerate) their disparate business units. Internal agglomeration will be higher between business units that have higher resource relatedness.

However, we know from existing studies that unrelated firms can also achieve synergies in resource use, and that these tend to be localized (Rosenthal and Strange 2003). Will the agglomerative force be stronger inside the firm or among unrelated firms? We expect that the tendency to agglomerate internally will be higher than the agglomeration of unrelated firms. While internal agglomeration can bring about benefits through shared resource use, it does pose the negative externalities of unrelated agglomeration. When unrelated firms decide to agglomerate, they also face costs from doing so, for example, price competition (Baum and Haveman 1997), competition for factor inputs and financial resources, congestion, and undesirable spillovers of knowhow to competitors (Alcácer 2006). Shaver and Flyer (2000) show, for example, that larger firms locate *away* from clusters because they want to limit such negative externalities and potential spillovers of their unique resources to competitors. Therefore, we propose that:

Hypothesis 2: The internal agglomeration of multibusiness firm will be higher than the agglomeration that we observe between unrelated firms.

Finally, which kinds of resources uses will predict the highest degree of internal agglomeration? As the examples of the capital machinery and brand made obvious, the degree of spatial transactions costs differs for different resources. On the one extreme, slack in a specialized item of capital equipment should predict high levels of proximity. On the other extreme, the use of a firm's brand is not highly sensitive to geographic distance. The spatial transaction costs related to the sharing and reallocation of goods and labor inputs likely lie somewhere between these extreme cases. Managers and workers are mobile, but the time spent commuting between different units of the firm present a rival use of the resource. On the other hand, goods face transportation costs, though likely lower than human resources. Moreover, recent findings put into question how important internal resource flows are even in vertically integrated firms (Atalay, Hortaçsu and Syverson 2014). In this study, we collect data on industries' relatedness in labor, input-output linkages, and technology linkages. Among those, we propose that:

Hypothesis 3: Industry relatedness in labor will be the strongest predictor of internal agglomeration. Input-output linkages and technological linkages between industries will be less significant predictors of internal agglomeration.

2.3 Data and Methodology

The data used in the study come from Dun & Bradstreet (D&B), a private company that for more than 100 years, has maintained and provided directories and commercial data on businesses. Prior studies using the data (e.g. Alfaro and Charlton 2009) and have shown that for U.S. firms, it closely tracks key statistics reported by the U.S. Census. While D&B maintains a global dataset, due to the long tradition of data collection and well-developed systems in the U.S., the U.S. data come close to a census of U.S. business activity. In this paper, we use the 2012 vintage of the data.

We create our sample by selecting all establishments located in the continental U.S. with U.S.-based headquarters. We take out establishments of firms headquartered elsewhere as other factors, such as proximity to the foreign headquarter, may be driving their location choices. Because the D&B data report business-owners as employees (unlike the Census) and thus over-report employment in small firms, we limit our sample to firms with five employees or more. From this set of firms, we select all firms that report any production in manufacturing, i.e. an SIC three-digit code in the 201–399 range.[3] We limit our analysis to manufacturing for several reasons. One is to facilitate the comparability of our results with prior studies, especially the literature on external agglomeration, which has almost exclusively been limited to manufacturing (e.g Shaver and Flyer 2000, Ellison, Glaeser and Kerr 2010). The second reason is driven by the availability of proxies for industry relatedness, for example input-output linkages, which are much more difficult to meaningfully construct for non-manufacturing industries. Finally, due to the computational intensity involved in generating continuous agglomeration measures, which are fed millions of distance observations, expanding the sample of industries would exponentially increase the number of possible industry pairs and the cost of calculating the needed distance measures for each industry dyad.

---

[3]Following prior studies, we exclude ten industries from the analysis which are unusual and sparse in the data and these are tobacco (211–214), fur goods (237), portions of printing and publishing (277–279), secondary nonferrous metals (334), and search and navigation equipment (381) (henceforth, “excluded industries").

After applying the sample selection criteria above (i.e. continental-U.S. plants in manufacturing with five or more employees) we are left with observations on 269,743 distinct firms and their 327,895 manufacturing establishments (Table 2.1). 106,953 (33 percent) of the establishments belong to 48,801 distinct multibusiness firms while the remainder are stand-alone, producing in only one manufacturing industry and one location. Note that for each establishment, D&B reports up to six different industry codes. Table 2.1 shows that among the multibusiness firms in the data, the majority are firms that that produce in only one location (single-location multibusiness firms). In robustness checks, we will exclude all observations that represent co-location within an establishment to ensure that the results do not depend on such immediate co-location, but are a more general feature of the location patterns of different establishments belonging to multibusiness firms.

Table 2.1: *Firms and Establishments Included in the Analysis*

	Firms		Establishments		Workers (in 1,000s)	
Total multibusiness, of which	48,801	18%	106,953	33%	9,845	66%
Multi-location	12,901		71,053		8,556	
Single location	35,900		35,900		1,290	
Stand-alone	220,942	82%	220,942	67%	5,085	34%
Total	269,743		327,895		14,930	

Note: Data from Dun & Bradstreet, 2012. Sample includes all establishments located in the continental U.S. with more than five workers reporting a product code that falls in SIC3 industries 201–399 (with the exception of ten excluded industries). Employment data are from the same source.

2.3.1 Measuring Agglomeration in Multibusiness Firms

Our main outcome of interest is the internal agglomeration of business units within multibusiness firms. Two main approaches to measuring the tendency of industries to locate in geographical proximity have emerged. The first approach summarizes to what extent firms are co-located in the same discrete spatial unit, for example a state, a county, or a ZIP code. One of the most widely used discrete measures is the Ellison-Glaeser index of agglomeration

(Ellison and Glaeser 1997), which calculates the share of manufacturing employment in the specified industries in a location, relative to the amount of total manufacturing employment in the location.

However, discrete indexes suffer from several limitations. One is that they treat all spatial units symmetrically, so that plants in neighboring units are treated the same way as plants far away. This is a potential source of downward bias when measuring agglomeration, as firms co-located across the spatial unit boundaries, even if geographically close, will be assigned a zero value of co-location. A second criticism is that aggregating individual observations to larger spatial units can results in the "modifiable areal unit problem" whereby the choice of boundaries can have a large effect on the underlying concepts measured and the aggregation of individual data to larger units can create artificial correlations and bias.

We proceed by measuring agglomeration using a continuous distance measure, but also provide secondary results using discrete spatial units. The dominant continuous measure is the Duranton-Overman index (Duranton and Overman 2005) which estimates the likelihood that two establishments are located at any given distance d with a kernel density estimator. Specifically, for every industry pair i and j and distance d, that likelihood is estimated using:

$$\hat{K}_{ij}(d) = \frac{1}{n(n-1)h} \sum_{r=1}^{n_i} \sum_{n=1}^{n_j} f\left(\frac{d - d_{r,s}}{h}\right) \tag{2.1}$$

where $\hat{K}_{ij}(d)$ is the kernel density at distance d for industry-pair ij, $d_{r,s}$ is the straight-line distance between establishments r and s in industries i and j,[4] n is the number of establishment observations, f is the Gaussian kernel function, and h is the bandwidth that minimizes the mean squared error. The measure answers: what is the likelihood that two establishments picked at random, one in industry i and the other in j are located d kilometers apart? The likelihood is estimated at every one-kilometer increment from 1 to more than 4,000 kilometers (the maximum distance in the continental U.S.). We also estimate

[4]We calculate distance between establishments using the coordinates of the centroids of their ZIP codes and the straight-line distance between them per the Haversine formula. Geo-coordinates of ZIP codes are obtained from http://www.unitedstateszipcodes.org/zip-code-database/. We use 6,371 kilometers as the measure of the Earth's radius in the Haversine formula.

an employment-based likelihood function, where each observation is a distance between an employee pair in establishments r and s, rather than the establishments themselves. Thus, the employment-based measure gives more weight to larger establishments in the calculation of the overall agglomeration densities.

To assess whether agglomeration is high or not, one needs a benchmark against which to compare the estimated likelihood. In agglomeration studies, this benchmark is usually a measure of "random" agglomeration, which would be the degree of agglomeration observed if all firms were randomly picking locations from the set of observed sites. Operationally, researchers calculate the index once using the actual establishment locations and then again by replacing the actual locations of establishments with random draws of the locations observed in the data. In this paper, we make the comparison not between overall and random agglomeration but rather between internal and external agglomeration (keeping in mind that existing research finds that observed external agglomeration tends to be larger than random). Thus, we calculate the Duranton-Overman index once for each industry pair using all the locations of plants in the sample of multibusiness firms and then again using plant locations from the sample of stand-alone firms.[5]

While the approach in each sample is parallel, one difference relates to the underlying number of observations over which the distance measures are taken. In the case of internal agglomeration, we can construct a population measure, taking into account all the instances when industry i and j are observed *in the same firm*, and recording the distance between the plants.[6] In the case of stand-alone firms, however, forming all possible dyads between plants in industry i and j would result in an extremely large number of observations, making computation close to impossible. Therefore, we follow the standard approach in

[5]The Appendix illustrates one example using industry pair SIC 282 and SIC 291 (plastics materials and petroleum refining). Figure A.11 shows the locations for all plants of stand-alone firms in these two industries and Figure A.12 shows the plants in these two industries belonging to one selected multibusiness firm, Chevron.

[6]Note that if a multibusiness firm reports more than one industry in the same establishment, we take the geographic distance between these industries for that establishment to be zero. As a robustness check, we recalculate a measure of internal agglomeration excluding these within-establishment observations in order to ensure that the results are not sensitive to their exclusion.

the literature and, for each industry i and j select 1,000 firms at random and take all of the distance measures between those plants. In the cases where an industry is associated with less than 1,000 establishments in the data, we select all of the establishments.[7]

Measured in this way, each measure of agglomeration of industry pairs i and j is based on up to 1,000,000 observations of distance in the stand-alone sample and is based on the actual number of times two industries are observed within a firm in the multibusiness sample. As that number is generally smaller than 1,000,000, we expect that the error bounds on the estimates of internal agglomeration to be larger. Indeed, there are some industry pairs which are rarely or never observed within the same firm, for example Men's and Boy's Suits and Coats (SIC 231) and Metalworking Machinery (SIC 354). To address this small-N problem, throughout the study we present baseline results using only those industry pairs which are observed within the same firm at least 30 times. This reduces the number of industry pairs from 8,385 possible pairs to 3,973 pairs.

2.3.2 Explanatory Variables

Input-output (IO) linkages: Following prior studies (e.g Fan and Lang 2000, Lemelin 1982), we measure input-output linkages using data form the Benchmark Input-Output Accounts published by the U.S. Bureau of Economic Analysis (BEA).[8] For each industry i we calculate $InputRequirement_{ij}$, the share of i's inputs that originate with industry j.[9] We also calculate the share of industry i's output purchased by industry j, $OutputShare_{ij}$. The measure of

[7]Plant counts per industry range from under 100 to more than 41,000 at the SIC3 digit level in the D&B data. 40 industries in the data have fewer than 1,000 stand-alone establishments.

[8]We use the 1992 tables, which are the most recent ones for which the BEA provides a concordance from the six-digit input-output industry classification to the 1987 SIC4 level industrial classification, which is used in the Dun & Bradstreet data. Ultimately, we aggregate to three-digit SIC industries because other data required for this study (e.g. technological input-output linkages) are available at that level of aggregation. We use the regularized IO-SIC concordance provided by Davin Chor. The concordance maps 361 IO industries to 459 SIC4 manufacturing industries. As a result, while each SIC manufacturing industry maps to only one IO industry, some IO industries map to multiple SIC industries. In these cases, we assign the IO industry's output to SIC4 industries in two different ways: 1) proportionally to the SIC4 industry's total shipments in 1992 and 2) equally across each of the SIC industries. While we use the shipment-weighted measure in our main analysis, the correlation between the two weights is 0.959 and the results are very similar using either measure.

[9]Including manufacturing and non-manufacturing inputs, except wholesale and retail trade.

input-output linkages, then is the maximum of the input and output share. Since these measures are asymmetric for industry pair ij, we take again the maximum at the industry-pair level to create a symmetric measure of the maximum input-output linkage in industry pair ij. Table 2.2 shows summary statistics of the measure. While the average input-output linkages among two industries of 1.2 percent is quite low, some industries are strongly vertically connected, with a maximum input-output linkage of 61.5 percent.

Labor similarity: To measure how similar industries are in the labor that they employ we use the 2001 industry-occupation matrix from the U.S. Bureau of Labor Statistics (BLS). The matrix shows for each SIC3 industry the number of workers employed in each of 96 different occupation categories.[10] Labor similarity measures the correlation between the vectors of occupational shares. Table 2.2 shows the summary statistics. While the average labor similarity of industry pairs is 0.50, the maximum is close to 1.0 (this is the labor similarity for Men's and Boys' Furnishings, Work Clothing, and Allied Garments (232) and Women's, Misses', and Juniors' Outerwear (233).

Technological similarity: To measure the tendency of two industries to share relevant knowledge, we follow the procure used in calculating the input-output linkages but replacing the source data with data of patent citations between industries i and j. The underlying patent data are mapped to SIC3 industries using the concordances produced by Brian Silverman (1999).[11] Specifically, we calculate $PatentIn_{ij}$ as the share of patents cited by industry i that originate in industries i though j and $PatentOut_{ij}$ as the share of industry i's patents cited by industries i through j. Then the $TechLink_{ij}$ measure is the maximum of the pairwise measures.

Given that we are estimating the relationship between industry-pair relatedness and the overall tendency of two industries to agglomerate, the estimation is the level of industry

[10]The BLS publishes the data at four different levels of aggregation: 1) major group (23 occupations); 2) minor group (96 occupations); 3) broad occupation (449 occupations); and 4) detailed occupation (821 occupations). Given the number of SIC3 industries is 130, the 96 minor groups result in the most appropriate mapping (finer occupation categories, e.g. 449 occupations, would have a number of occupations that map only to one or two industries). http://www.bls.gov/soc/2000/socguide.htm.

[11]We use the publically available data provided in the data annex of Ellison, Glaeser and Kerr (2010).

pair dyads. We estimate a model of the following form, separately for the sample generated from data on multibusiness firms and from the data of stand-alone firms:

$$Agg^S_{ij,d} = \alpha + \beta_1 IO\,links_{ij} + \beta_2 LaborSim_{ij} + \beta_3 Tech\,links_{ij} + \gamma_i + \gamma_j + \varepsilon_{ij} \;\; S \in \{MB, SA\} \tag{2.2}$$

where $Agg_{ij,d}$ is the agglomeration (the cumulative density) of industries i and j at a distance of d kilometers. In general, we present results at the 10, 50, 250, and 500 kilometer thresholds, though other choices yield similar patterns. We normalize the dependent and independent variables, in order to facilitate comparability among the effects of the different drivers of agglomeration. The estimated coefficients β_1, β_2, and β_3 represent the marginal effects of a one standard deviation increase in industry relatedness on agglomeration. Throughout the analysis, we control for fixed effects for each industry (the γ), to account for the fact that some industries are simply more concentrated (disbursed) than others and thus may have low (high) agglomeration with any other industry. As agglomeration is a generated regressor (in the stand-alone sample), we report bootstrapped standard errors.

2.4 Results

2.4.1 The Level of Internal Agglomeration

Table 2.2 shows the summary statistics of the key variables. Comparing the level of agglomeration for multibusiness versus stand-alone firms we see that at all distance thresholds, the average and median internal agglomeration is higher, and in general, roughly twice as high as external agglomeration (Appendix Figure A.10). For example, the cumulative probability that two randomly picked establishments are at a distance of 250 km or less is 10.2 percent when they are inside the same firm, versus 5.7 percent when they are stand-alone firms.

Table 2.2: *Summary Statistics*

	Obs	Mean	St. Dev.	Median	Min	Max
Share of plants:	Internal agglomeration - Firm count-based					
at d $\leq$ 10 km	3736	0.004	0.004	0.004	0.000	0.119
at d $\leq$ 50 km	3736	0.020	0.017	0.018	0.001	0.511
at d $\leq$ 250 km	3736	0.102	0.057	0.092	0.005	0.872
at d $\leq$ 500 km	3736	0.208	0.091	0.193	0.016	0.961
in the same county	3736	0.008	0.013	0.004	0.000	0.261
in the same CBSA	3736	0.021	0.028	0.014	0.000	0.490
in the same state	3736	0.066	0.061	0.053	0.000	0.878
Share of plants:	Internal agglomeration - Employment count-based					
at d $\leq$ 10 km	3736	0.003	0.005	0.002	0.000	0.114
at d $\leq$ 50 km	3736	0.018	0.026	0.012	0.000	0.670
at d $\leq$ 250 km	3736	0.088	0.078	0.071	0.000	0.953
at d $\leq$ 500 km	3736	0.196	0.121	0.174	0.000	0.955
in the same county	3736	0.008	0.019	0.002	0.000	0.397
in the same CBSA	3736	0.021	0.036	0.011	0.000	0.596
in the same state	3736	0.064	0.072	0.046	0.000	0.953
Share of plants:	External agglomeration - Firm count-based					
at d $\leq$ 10 km	3736	0.002	0.001	0.002	0.001	0.008
at d $\leq$ 50 km	3736	0.011	0.003	0.011	0.003	0.037
at d $\leq$ 250 km	3736	0.057	0.013	0.056	0.024	0.147
at d $\leq$ 500 km	3736	0.126	0.018	0.124	0.077	0.231
in the same county	3736	0.006	0.004	0.005	0.001	0.060
in the same CBSA	3736	0.013	0.007	0.011	0.001	0.119
in the same state	3736	0.049	0.012	0.047	0.023	0.148
Share of plants:	External agglomeration - Employment count-based					
at d $\leq$ 10 km	3736	0.002	0.001	0.002	0.000	0.014
at d $\leq$ 50 km	3736	0.011	0.005	0.011	0.002	0.067
at d $\leq$ 250 km	3736	0.057	0.015	0.055	0.020	0.193
at d $\leq$ 500 km	3736	0.127	0.021	0.125	0.071	0.290
in the same county	3736	0.006	0.004	0.005	0.001	0.062
in the same CBSA	3736	0.013	0.007	0.011	0.002	0.122
in the same state	3736	0.047	0.012	0.045	0.018	0.155
Input-output linkages	3736	0.012	0.037	0.001	0.000	0.615
Labor similarity	3736	0.501	0.287	0.456	0.024	0.999
Technology linkages	3736	0.024	0.034	0.013	0.000	0.427

Note: Data from Dun & Bradstreet, 2012. Employment data are from the same source. Sample includes all establishments located in the continental U.S. with more than five workers reporting a product code that falls in SIC3 industries 201–399 (with the exception of ten excluded industries). 3,736 is the number of industry pairs in SIC industries 201–399 observed at least 30 times in a multi-business firm.

Table 2.3: *Results of T-Tests of Equality of Means*

At distance treshold	Two-sided T-test p-value H0: Multi-business = Stand-Alone	Two-sided T-test p-value H1: Multi-business > Stand-Alone
10 km	0.000	1.000
50 km	0.000	1.000
250 km	0.000	1.000
500 km	0.000	1.000

The table shows the results of a two-tailed t-test of the equality of means of the internal and external agglomeration density at each distance treshold. Agglomeration density is measured for each of 3,736 industry pairs.

Note that, as expected given the number of underlying distance observations, the standard deviation of the agglomeration estimate in multibusiness firms is also much larger than in stand-alone firms. In order to conclude that internal agglomeration is indeed higher than external agglomeration, we formally test the difference of the levels. We present a two-tailed t-test of the equality of means of internal and external agglomeration at different distance threshold in Table 2.3. In each case, the null hypothesis of equal means is rejected in favor of the conclusion that internal agglomeration is higher than external agglomeration.

Stylized fact 1: The internal agglomeration of multibusiness firms is higher than the external agglomeration of unrelated firms.

2.4.2 Drivers of Internal and External Agglomeration

Next, we introduce the determinants of agglomeration into the analysis in order to investigate the drivers of agglomeration in multibusiness and stand-alone firms. Table 2.4 presents the results of univariate regressions (each agglomeration determinant, one at a time) while Table 2.5 presents the results jointly controlling for all three determinants, per the model presented in Equation 2.2.

The results reveal some interesting patters. For internal agglomeration of multibusiness firms, labor similarity has the largest positive effects in predicting higher agglomeration. The effect size on labor similarity ranges from 0.06 to 0.11 meaning that a one standard deviation increase in labor similarity is associated with up to a 0.11 standard deviation increase in agglomeration — an elasticity of roughly one-tenth. Input-output linkages have the second largest effects, ranging from 0.02 to 0.04, followed by technological linkages with effect sizes from 0.02 to 0.03. Moreover, the coefficient on labor similarity is positive and significant at each distance threshold. Meanwhile the coefficients on input-output and technology linkages are significant only at the higher distance thresholds. This suggests that input-output and technology linkages don't predict that business units are very close, though they predict that they are not too far apart.

Table 2.4: *OLS results of Univariate Regressions of Internal and External Agglomeration on Industry Relatedness*

	Internal Agglomeration				**External Agglomeration**			
	10 km.	50 km.	250 km.	500 km.	10 km.	50 km.	250 km.	500 km.
IO linkages	0.022	0.022	0.036**	0.044***	0.056***	0.056***	0.064***	0.069***
	(0.016)	(0.016)	(0.017)	(0.016)	(0.009)	(0.009)	(0.011)	(0.012)
R^2	0.346	0.357	0.498	0.492	0.017	0.018	0.001	0.511
Labor similarity	0.055**	0.056*	0.098***	0.107***	0.115***	0.116***	0.155***	0.196***
	(0.028)	(0.029)	(0.026)	(0.026)	(0.010)	(0.010)	(0.011)	(0.014)
R^2	0.347	0.357	0.499	0.494	0.915	0.914	0.892	0.834
Tech. linkages	0.018	0.020	0.033***	0.031**	0.046***	0.046***	0.054***	0.063***
	(0.013)	(0.013)	(0.012)	(0.012)	(0.005)	(0.005)	(0.006)	(0.008)
R^2	0.346	0.356	0.497	0.491	0.912	0.912	0.887	0.826
Observations	3,736	3,736	3,736	3,736	3,736	3,736	3,736	3,736
Each industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each cell reports the results of a univariate regression of internal (external) agglomeration on a measure of industry relatedness. Agglomeration is calculated separately for multibusiness and stand-alone firm establishments using 2012 data on all U.S. manufacturing establishments with more than five employees from Dun & Bradstreet. We only include industry pairs observed at least 30 times within a firm, resulting in 3,736 industry-pair observations. Input-output linkages are calculated with 1992 data from the BEA benchmark input-output table. Labor similarity is calculated with the 2001 national industry-occupation employment matrix from the BLS. Technology linkages are based on the NBER's patent citation database from 1975-1997. Each regression includes fixed effects for each industry. All variables are normalized. Bootstrapped standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.5: *OLS results of Multivariate Regressions of Internal and External Agglomeration on Industry Relatedness*

	Internal Agglomeration				**External Agglomeration**			
	10 km.	50 km.	250 km.	500 km.	10 km.	50 km.	250 km.	500 km.
IO linkages	0.016	0.015	0.024	0.032**	0.042***	0.042***	0.046***	0.046***
	(0.017)	(0.016)	(0.017)	(0.016)	(0.009)	(0.009)	(0.011)	(0.012)
Labor similarity	0.049*	0.049*	0.088***	0.096***	0.097***	0.099***	0.136***	0.176***
	(0.028)	(0.029)	(0.026)	(0.026)	(0.009)	(0.009)	(0.011)	(0.014)
Tech. linkages	0.007	0.009	0.015	0.009	0.022***	0.022***	0.023***	0.027***
	(0.014)	(0.014)	(0.012)	(0.013)	(0.006)	(0.006)	(0.007)	(0.009)
Observations	3,736	3,736	3,736	3,736	3,736	3,736	3,736	3,736
R^2	0.347	0.357	0.500	0.495	0.917	0.916	0.894	0.837
Each industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each column reports the results of a multivariate regression of pairwise agglomeration on all three agglomeration determinants. Agglomeration is calculated separately for multibusiness and stand-alone firm establishments using 2012 data on all U.S. manufacturing establishments with more than five employees from Dun & Bradstreet. We only include industry pairs observed at least 30 times within a firm, resulting in 3,736 industry-pair observations. Input-output linkages are calculated with 1992 data from the BEA benchmark input-output table. Labor similarity is calculated with the 2001 national industry-occupation employment matrix from the BLS. Technology linkages are based on the NBER's patent citation database from 1975-1997. Each regression includes fixed effects for each industry. All variables are normalized. Bootstrapped standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Turning to the drivers of external agglomeration in the sample of stand-alone firms, all determinants are found to be individually significant at every distance threshold. This is in line with prior findings (Ellison, Glaeser and Kerr 2010) and confirms that the existing conclusions of the literature are not sensitive to omitting the external agglomeration of multi-business firms. In terms of magnitude, here also labor similarity has the largest effects ranging from 0.12 to 0.20, followed by input-output linkages with effect sizes in the range of 0.06 to 0.07, and finally technological linkages with effects in the 0.05 to 0.06 range.[12]

[12]These effects are generally in line with those reported by Ellison, Glaeser and Kerr (2010), though somewhat smaller in particular for input-output linkages. This can be due to different industry samples over which the estimation is calculated or differences in the underlying plant location patterns, which may have evolved (Ellison, Glaeser and Kerr (2010) is based on the 1987 and 1997 Census of Manufacturers). Falling transportation costs and increased international outsourcing since that period would be consistent with the lesser role of input-output linkages which we observe when calculating agglomeration with the 2012 Dun & Bradstreet data.

The results of the multivariate regressions, where all three measures of industry relatedness are included in the model together, presented in Table 2.5 lead to similar conclusions. While all agglomeration determinants are still separately significant in the sample of stand-alone firms, in the sample of multibusiness firms, only labor similarities are significant at all levels of distance, input-output linkages are significant only in predicting agglomeration of 500 kilometers and less and technology linkages are small in effect size and not significant at any threshold of distance. The results of these analyses are summarized in the following:

Stylized fact 2: The drivers of agglomeration among multibusiness and stand-alone establishments are somewhat distinct. Labor similarity is the largest and most significant predictor of internal agglomeration. Input output linkages predict that plants are not too far away. Technology linkages are not a significant predictor of internal agglomeration. All three determinants predict external agglomeration at all distance thresholds.

2.4.3 Differences of Internal and External Agglomeration

Finally, following the approach of Alfaro and Chen (2014), we examine the difference in internal and external agglomeration levels and test whether the measures of industry relatedness are able to explain the tendency of multibusiness firms to agglomerate more than stand-alone firms. The estimating equation is as in Equation (1), with the dependent variable replaced by the *difference* in the agglomeration levels of multibusiness and stand-alone firms for each industry pair ij, specifically $Agg^{MB}_{ij,d}$-$Agg^{SA}_{ij,d}$. The results are presented in Table 2.6. They point to the importance of labor similarities in driving the higher internal agglomeration of multibusiness firms. Specifically, a one standard deviation higher labor similarity of industry pair ij is associated with a 0.003–0.008 increase in the difference in the agglomeration density between multibusiness and stand-alone firms (where the mean difference ranges from 0.001–0.082 from the 10 to the 500-kilometer distance range). This evidence is consistent with a view that sharing a labor pool plays a more important role for the location decisions of multibusiness firms relative to stand-alone firms. Input-output linkages predict the higher agglomeration of multibusiness establishments at distances beyond 500 kilometers. Meanwhile, the coefficients on the technological linkages between industries, while positive, are not a statistically significant predictor of the higher agglomeration density of multibusiness firms.

Table 2.6: *OLS results of the Difference Between Internal and External Agglomeration*

	Internal-External Difference			
	50 km.	250 km.	500 km.	1000 km.
IO linkages	0.000	0.001	0.002	0.005***
	(0.000)	(0.001)	(0.001)	(0.002)
Labor similarity	0.000	0.003**	0.006**	0.008**
	(0.000)	(0.001)	(0.002)	(0.003)
Tech. linkages	0.000	0.001	0.000	-0.001
	(0.000)	(0.001)	(0.001)	(0.002)
Observations	3,736	3,736	3,736	3,736
R^2	0.343	0.472	0.464	0.424
Each industry FE	Yes	Yes	Yes	Yes

Note: Each column reports the results of a multivariate regression, where the dependent variable is the difference in agglomeration of multibusiness and stand-alone firm establishments, each calculated separately using 2012 data on all U.S. manufacturing establishments with more than three employees from Dun and Bradstreet. We only include industry pairs for which at least 30 observations of multibusiness firms are available, resulting in 3,973 industry-pair observations. Input-output linkages are calculated with 1992 data from the BEA benchmark input-output table. Labor similarity is calculated with the 2001 national industry-occupation employment matrix from the BLS. Technology linkages are from EGK based on the NBER's patent citation database from 1975-1997. Each regression includes industry fixed effects. All variables are normalized, except the dependent variable which is in levels. Bootstrapped standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

2.5 Further Tests and Robustness Checks

Table 2.7 presents the results of several additional analyses of internal agglomeration and its drivers. For the first robustness check, we re-calculate internal agglomeration but excluding all observations of industries manufactured in the same plant, with a distance equal to zero. We saw per Table 2.1 that a meaningful number of multibusiness firms are single-plant multi business firms, manufacturing more than one industry in the same plant. This suggests that these activities have a high amount of resource interdependence. However, we want to ensure that this extreme type of co-location is not driving the overall result that we observe. The results excluding sample-plant agglomeration are presented in columns (1) and (2) of Table 2.7 for two different distance thresholds. The pattern of results is unchanged from the main analysis.

Next, we present results from an analysis where the count-based measure of agglomeration is replaced with an employment-weighted measure. The count-based measure weights each establishment equally in calculating the distribution of distances between establishments. The employment-based measure, on the other hand, weighs each establishment by its employment. The employment-based measure effectively answers: what is the likelihood that two *workers* one in industry i and the other in j inside a multibusiness firm picked at random, are located a distance d apart? If it is, for example, the case that firms co-locate their larger plants, then the employment based measure would be larger than the count-based measure. If smaller firms tend to co-locate more than large ones, than the employment-based measure would be smaller than the count-based one. From Table 2.7 we can observe that for our multibusiness firm sample, the employment-based agglomeration measure tends to be lower than the count-based, suggesting that inside the firm, smaller plants tend to agglomerate more. For the stand-alone sample, on the other hand, the employment-based and the count-based measures show very little difference.

The results using the employment-based measure are presented in columns 3 and 4 of Table 2.7. First, note that the R^2 is lower, i.e. the distance between large plants inside a firm is harder to predict with industry fixed effects and the measures of industry relatedness than the distance of smaller plants. Second, in terms of the drivers, all three measures of industry relatedness are now positive and statistically significant, both at low (Column 3) and higher distance thresholds (Column 4). This finding suggests that among the bigger plants inside a firm, all three drivers of industry relatedness predict agglomeration. For example, if two establishments are in vertically related industries *and* they are large, then they are likely to be close together inside the firm. This suggests that the main results where linkages matter less are driven the behavior of the smaller plants inside the firm which do not agglomerate as strongly based on these linkages.

Finally, we also present results using discrete spatial units rather than the continuous agglomeration index. The dependent variable now is the share of plants within a firm in industries i and j that are in the same county, core-based statistical area (CBSA), and state. The results on labor similarity of these regressions are very similar to the main results. With the discrete spatial units, the results on input-output and technology linkages are somewhat different, with input-output linkages associated with higher agglomeration at the county and CBSA but not the state level. Meanwhile technology linkages are associated with greater agglomeration at the county and state levels.

2.6 Discussion and Conclusion

This study has presented evidence for the view that the potential for the sharing of resources is an important driver of the location decision of the different units of multibusiness firms. Labor similarities, in particular, exert an important centripetal influence on multibusiness firms. Input-output similarities appear to matter less for keeping firm units close together, but exert some influence at higher distance thresholds, keeping activities not too far apart. We have weak evidence that the shared use of technological inputs (proxied by patens) exert a distinct force in keeping multibusiness firm units geographically proximate.

The findings that multibusiness firms locate so as to leverage internal resources contribute novel insights in both the corporate strategy and economic geography literatures. In general, it is believed that external agglomeration economies, benefits steming from the co-location between unrelated firms, boosts firm productivity (e.g. Martin, Mayer and Mayneris 2011). To the extent that multibusiness firms can generate similar productivity advantages internally, through the co-location of their business units, this provides a valuable and rare resource that cannot be imitated by single-business firms. Thus, their location strategies may be an important way in multibusiness firms can achieve better performance outcomes relative to single-business firms.

This work also raises further questions. One somewhat surprising result is that technology linkages, which proxy knowledge exchange between business units, appear to play little role in internal agglomeration. Besides potential weaknesses of this variable in actually measuring the exchange of knowledge inside a firm, this finding could be consistent with at least two other explanations. One is that while multibusiness firms locate their establishments so as to facilitate the movement of goods and workers, they care little about facilitating the transfer of know-how. A second, and we believe more plausible explanation based on existing evidence from prior studies is that multibusiness firms share knowledge internally between their establishments but that the "transportation costs" of these knowledge flows are lower *within* than *between* firms (Almeida, Song and Grant 2002). The ability of multibusiness firms to transfer knowledge even at large distances thus reduces the incentive to collocate in order to facilitate these flows. While the coarseness of our current data prevent us from doing so in this project, being able to more clearly disentangle which of these mechanisms is at play constitutes an important research question.

A second interesting finding points to a potentially important role of human capital resources and thus internal labor markets as a mechanism though which multibusiness firms can achieve synergies and reallocate resources. This is an interesting area of further inquiry. Recently, a handful of papers have documented that multibusiness firms appear to use internal labor markets to adjust to shocks in some business units (Tate and Yang 2015) as well as to avoids the frictions in hiring and firing in external labor markets (Belenzon and Tsolmon 2016). A separate possibility, however, is that internal labor markets themselves serve as conduits of knowledge inside the firm. Gaining more insight into when, how, and why multibusiness firms redeploy workers internally is another promising area for future research.

Table 2.7: *Robustness Checks*

	No within-estab.		Employment-weighted		Discrete spatial units		
Dependent variable:	*Agg. at 50km*	*Agg. at 250km*	*Agg. at 50km*	*Agg. at 250km*	*Same county*	*Same CBSA*	*Same state*
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
IO linkages	0.015	0.024	0.065***	0.068***	0.038**	0.068***	0.034
	(0.016)	(0.017)	(0.022)	(0.021)	(0.019)	(0.018)	(0.021)
Labor similarity	0.050*	0.088***	0.056**	0.098***	0.081***	0.088***	0.100***
	(0.029)	(0.026)	(0.025)	(0.025)	(0.027)	(0.029)	(0.028)
Tech. linkages	0.009	0.015	0.046**	0.048***	0.064***	0.029	0.065***
	(0.014)	(0.012)	(0.022)	(0.017)	(0.022)	(0.018)	(0.018)
Observations	3,736	3,736	3,736	3,736	3,736	3,736	3,736
R^2	0.357	0.500	0.251	0.364	0.180	0.246	0.324
Each industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each column reports the results of a multivariate regression of pairwise agglomeration in multibusiness fimrs on all three agglomeration determinants. Note: Columns (1) and (2) exclude industries co-located in the same establishment in the calculating agglomeration. Columns (3) and (4) weigh each establishment by its employment (as reported by Bun & Bradstreet) in calculating agglomeration. Columns (5)-(7) calculate the percentage of times that establishments in industry pair *ij* are observed in the same county, core based statistical area (CBSA) or state. We only include industry pairs observed at least 30 times within a firm, resulting in 3,736 industry-pair observations. Input-output linkages are calculated with 1992 data from the BEA benchmark input-output table. Labor similarity is calculated with the 2001 national industry-occupation employment matrix from the BLS. Technology linkages are based on the NBER's patent citation database from 1975-1997. Each regression includes fixed effects for each industry. All variables are normalized. Bootstrapped standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.8: *Correlations of Main Variables*

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Internal agg. at 10 km	1.000																
2	Internal agg. at 50 km	0.997	1.000															
3	Internal agg. at 250 km	0.880	0.895	1.000														
4	Internal agg. at 500 km	0.766	0.784	0.964	1.000													
5	Internal - % in the same county	0.257	0.267	0.307	0.263	1.000												
6	Internal - % in the same CBSA	0.506	0.504	0.475	0.431	0.504	1.000											
7	Internal - % in the same state	0.615	0.619	0.699	0.640	0.422	0.510	1.000										
8	External agg. at 10 km	0.152	0.156	0.201	0.204	0.077	0.131	0.106	1.000									
9	External agg. at 50 km	0.153	0.157	0.203	0.206	0.077	0.132	0.107	1.000	1.000								
10	External agg. at 250 km	0.169	0.173	0.226	0.231	0.076	0.147	0.120	0.975	0.977	1.000							
11	External agg. at 500 km	0.164	0.170	0.226	0.232	0.073	0.158	0.127	0.757	0.762	0.876	1.000						
12	External - % in the same country	0.150	0.151	0.178	0.162	0.058	0.129	0.123	0.821	0.820	0.768	0.495	1.000					
13	External - % in the same CBSA	0.127	0.130	0.149	0.140	0.002	0.139	0.107	0.462	0.463	0.462	0.340	0.625	1.000				
14	External - % in the same state	0.094	0.095	0.117	0.104	0.085	0.094	0.097	0.778	0.776	0.699	0.395	0.864	0.415	1.000			
15	Input-output linkages	0.114	0.116	0.153	0.158	0.073	0.129	0.147	0.097	0.098	0.120	0.152	0.097	0.164	0.073	1.000		
16	Labor similarity	0.114	0.117	0.162	0.154	0.119	0.111	0.172	0.089	0.090	0.110	0.135	0.117	0.031	0.125	0.238	1.000	
17	Technology linkages	0.018	0.019	0.034	0.033	0.108	0.069	0.104	0.019	0.019	0.016	0.028	0.019	-0.006	0.072	0.236	0.263	1.000

Chapter 3

Internal Labor Markets in Multibusiness Firms

3.1 Introduction

[1]In recent years, an active stream of research has developed around the theory of resource redeployment, the view that firms can generate excess value by actively managing their resources, withdrawing them from some business units and reallocating them to others in response to changing conditions (Helfat and Eisenhardt 2004, Levinthal and Wu 2010, Sakhartov and Folta 2014, 2015, Lieberman *et al.* 2016). The theory is attractive because it is able to explain a potential source of competitive advantage in diversified firms as well as diversification decisions.

Despite these theoretical advances, with the exception of studies on internal capital markets, we still have little evidence regarding how firms manage their internal resource pools. Even simple descriptive statistics for the prevalence of resource redeployment in multibusiness firms are needed (Folta, Helfat and Karim 2016). We also lack direct tests for key features of theory, in particular, studies showing which resources are redeployed and how firms' organizational features enable or constrain redeployment. One key challenge for work in this area is the paucity of internal firm data showing how resources are redeployed.

In this paper, we leverage a rich dataset to study how firms allocate one key resource — their human capital — though internal labor markets. As production processes have become

[1]This chapter is joint work with Christopher Poliquin, Harvard Business School.

more skill- and service-driven, human capital is a critical resource for many firms. Although a rich literature exists on external labor markets (e.g. reviewed in Mawdsley and Somaya (2016)), we know much less about how human capital is allocated and reallocated within the firm though *internal* labor markets.

We develop a simple framework to predict when a firm with labor needs in a focal business unit will staff a position by redeploying a worker internally instead of hiring a worker in the external labor market. The resource redeployment literature has tended to assume that some feature of a resource makes it uneconomical to transact in external markets. Our framework allows us to be precise about when and why firms would want to reallocate workers internally versus source externally. Specifically, we model the possibility that internal workers are distinct from workers available in external labor markets because they possess firm-specific human capital. In addition, we also model the costs of using external labor markets, such as the costs of hiring and firing.

By incorporating these realistic features of internal and external labor markets, the framework shows that two distinct types of motivations can drive internal redeployments. One is external labor market frictions, e.g. the costs of hiring new workers and the costs of firing existing workers. Even if internal and external workers were otherwise homogeneous, such frictions would create incentives for firms to redeploy workers internally in order to avoid these transactions costs and institutional voids (Khanna and Palepu 2000).

A theoretically different possibility is provided by the view that workers are resources, embodiments of knowledge (Kogut and Zander 1992, Grant 1996). Some of that knowledge may be non-codifiable (Teece 1981) and firm-specific (Barney 1991). Such knowledge is acquired over time and cannot be easily transferred outside the firm (Groysberg *et al.* 2008). In these cases, internal and external workers are not perfect substitutes. Even in a world lacking external labor market frictions, we would still observe redeployment motivated by the desire to allocate this firm-specific resource, embodied in the worker, to its most productive use within the firm.

Beyond these two distinct motivations for redeployment, our framework also incorporates the idea that a firm's organization — both its corporate strategy as reflected in the relatedness of its activities and its geography, i.e. the location of its business units — can enable or constrain the firm's ability to engage in worker redeployment. We model the relatedness of the firms' activities, in particular the occupational similarity of the firm's different industries, as an enabler of redeployment. Greater similarity increases the probability that the type of worker needed in a focal unit is actually available elsewhere in the firm. We model geographic distance between the origin and destination plants as an increase in the transfer costs involved in redeploying workers, for example relocation expenses and incentives paid to workers to encourage them to relocate.

Guided by the framework, we study the extent and drivers of worker redeployment leveraging a rich employer-employee matched dataset made available by the Government of Brazil, the *Relação Anual de Informações Sociais* (RAIS). In it, we can observe all workers employed at a firm, as well as their movements among the firm's plants. For the current analysis, we select a ten percent random sample of all multibusiness firms operating in Brazil from 2004–2014. During this time period, multi-business firms accounted for 14–17 percent of the total formal sector labor force of Brazil (Appendix Figure A.13).

In stylized facts, we observe that Brazilian multibusiness firms source a substantial share of their labor needs internally. On average 12.1 percent of workers hired in any year come from other establishments of the same firm. At any point in time, redeployed workers represent 5.5 percent of a plant's workers. Among workers leaving an establishment, 11.8 percent move to jobs within the same firm. This percentage is even higher when firms close an establishment; 21.8 percent of workers in establishments that are closing down move to new positions within the same firm.

We next analyze worker-level models to gain insight into which employees are more likely to be redeployed and thus infer the motivations of redeployment. Two findings emerge consistently. First, comparing otherwise similar workers employed at the same plant and occupation group in a year, workers with more firm-specific experience are more likely to be redeployed. All else equal, a worker with the average level of firm experience (2.9

years) has a 7.4 percent higher likelihood of being redeployed compared to a worker with no firm-specific experience. Second, within a given plant, workers higher in the organizational hierarchy are more likely to be redeployed. In particular, on average, 8.2 percent of an establishment's managers are redeployed in any year. This is nearly double the 4.4 percent of service and production workers that are redeployed. If, as commonly thought, valuable firm-specific human capital tends to reside with workers higher up in the hierarchy, these results are consistent with the hypothesis that firms use redeployment as a tool to reallocate valuable human capital resources.

In order to better tease apart a hypothesis of firm-specific human capital from other potential drivers of worker redeployment — for example, external labor market frictions (e.g. higher search and information frictions for managers) or redeployments due to workers' personal motivations — we test whether internally redeployed workers earn a wage premium over otherwise similar workers hired in the external labor market in the same plant, occupation, and year. We find strong evidence of wage premia to internally redeployed workers. Specifically, internally redeployed workers enjoy a nine percent higher contractual wage compared to otherwise similar workers hired into the same occupation and establishment though the external labor marker. Moreover, this premium is small for internal workers without firm-specific experience (i.e. workers hired and immediately redeployed) and rising steeply in a worker's years of firm-specific experience. These findings also are consistent with the existence of productivity-enhancing firm-specific human capital which allows internal workers to generate (and capture) excess returns. They are less consistent with pure frictions in external labor markets or moves motivated by workers' personal preferences.

Exploring to what extent large firing costs may be driving redeployments, we find that while plant exits are associated with more workers being redeployed, overall, only 11 percent of the redeployments that we observe are concurrent with plants shutting down. We do not find that plants shutting down become more likely to redeploy "blue-collar" workers who may otherwise present additional firing costs (e.g. due to unionization) (Cestone *et al.*

2017). Rather we find that when a plant exits, workers highest in the hierarchy and those with more firm-specific experience are more likely to be redeployed.

Finally, we explore how the firm's organizational features, in particular the relatedness of activities and its geographic footprint, affect the extent of worker redeployment. Here we model the volume of redeployments between all possible sets of origins and destinations (dyads) in a firm as a function of their industry relatedness, their geographic distance, and proxies for differences in the growth patterns of their respective industries. We find evidence consistent with the view that greater labor relatedness between industries facilitates the redeployment of workers while greater geographic distance between plants strongly discourages worker redeployment.

Taken together, the findings point to the conclusion that internal labor markets within multibusiness firms serve as a conduit though which firm-specific human capital is transferred among the firm's units. We find particularly strong evidence that firms redeploy their managerial human capital, and especially those workers with higher levels of firm-specific experience. We find strong evidence that workers with more firm-specific human capital earn excess rents in the form of higher wages.

The view of internal labor markets supported by our findings is quite distinct from other prevailing views. Until recently, much of the literature of internal labor markets focused on "vertical" labor markets, or "career ladders" — i.e. the processes through which workers move up the hierarchy within a given firm and the ways that firms can design appropriate promotion mechanisms for workers over their careers (e.g. Doeringer and Piore 1971). Recently, a literature has begun to emerge which, rather than *vertical* considers the unique aspects of *horizontal* internal labor markets, i.e. worker moves in multi-plant and multibusiness firms. Thus far, existing studies have focused primarily on the potential of internal labor markets to avoid frictions and rigidities of external labor markets (Belenzon and Tsolmon 2016) and to enable firms to adjust to unexpected shocks, e.g. by reallocating workers from plants that are shutting down to other parts of the firm (Tate and Yang 2015).

In this paper we propose that, beyond these possibilities, internal labor markets can also play the role of allocating valuable, firm-specific human capital to the parts of the firm where it is most needed. Our view is consistent with redeployment motivated by the existence of rare and valuable resources which are otherwise not available or easy to transact in external markets (Helfat and Eisenhardt 2004). Beyond offering evidence for this alternative motivation for internal labor market activity, our study is also unique in exploring the organizational enablers of worker redeployment. Our findings suggest that the relatedness between the different industries of the multi-business firm and the geographic proximity of the firm's units facilitate workers redeployment. This implies that firms for which worker redeployment is an important part of the strategy and a source of competitive advantage, face a trade-off between the objectives of expanding their geographic and product boundaries and facilitating the flows of workers though the firm's internal labor market.

3.2 Theory and Hypotheses

The resource-based view sees the firm as a collection of "those (tangible and intangible) assets which are tied semipermanently to the firm" (Wernerfelt 1984, p. 172). An important emphasis in this view is embodied in the word *tied*, which implies that these assets have features that make them difficult to transact in the open market. If assets are homogeneous or perfectly mobile, they are not a resource, which are those assets that are valuable, rare, imperfectly imitable, and not substitutable (Barney 1991).[2]

As a firm learns and grows, some resources get freed up and "slack" is created (Penrose 1959). Because slack resources are difficult to transact in the open market (Teece 1982), assuming that they are fungible across activities and that the firm cannot expand infinitely in its primary product, this provides an incentive the firm to diversify and thus gives rise to the multibusiness firm. Once diversified, the ability to generate synergies through the simultaneous use of resources across multiple activities provides economies of scope and is

[2]The resource-based view is one of serval theories of the firm. In alternative theories—for example, firms as a "nexus of contracts" (Fama 1980)—resource flows among the firm's units are not a necessary feature.

a source of competitive advantage for multibusiness firms.

More recent additions to this literature highlight that while some internal resources have the quality of being public goods within the firm (scale free) others are rival and their use in one part of the firm limits their use in other parts (non-scale free) (Levinthal and Wu 2010). For example, a firm's brand is a scale-free resource and can be simultaneously leveraged in different units of the firm. However, the time and skill of the firm's managers is a non-scale free resource. Although scale free resources lend themselves to the simultaneous use across business units and the generation of synergies, non-scale free resources do not.

However, firms can achieve competitive advantages in resource use even for resources that are non-scale free through resource redeployment (Helfat and Eisenhardt 2004). Rather than using resources simultaneously in two different business units, firms can redeploy resources from activities where there are less valuable and toward activities where they are more valuable. The benefits generated by a strategy of redeployment are termed inter-temporal economies of scope, as they generate competitive advantage through the ability to optimally adjust resource use across activities *over time*. Resource redeployment can help firms exit businesses with declining prospects while lowering the costs of starting or expanding operations in more promising areas (Lieberman, Lee and Folta 2016). Note that the original business unit does not necessarily close as part of a strategy of redeployment (Folta, Helfat and Karim 2016).

Although theoretically attractive, synergies and resource redeployment have been very difficult to study empirically. A key reason is the rarity of data internal to firms that show how they allocate resources among the different business units.[3] Notable are approaches based on the observations or resource reconfiguration within particular firms (Karim and Mitchell 2004), though those pose the question how generalizable strategies are across firms. Other approaches induce redeployment by observing business unit entries and exits (Lieberman *et al.* 2016), though the actual flows of resources are not observed.

[3] An important exception is the literature on internal capital markets, which has documented the extensive use of internal allocation mechanisms and the relative advantages of internal versus external capital markets, e.g. Lamont (1997), Stein (1997), Shin and Stulz (1998).

In this paper, we *directly* observe the movement and reallocation of one important type of resource across a large set of firms — workers. Human capital resources are one of the three resource categories identified by Barney (1991) and include "the training, experience, judgment, intelligence, relationships, and insight of *individual* managers and workers in a firm" (Barney 1991, p. 101, original emphasis). However, not all workers constitute resources. To the extent that a worker's attributes are homogeneous, and thus easily substitutable, then this worker would not be considered a resource. On the other hand, if a worker has some rare skills or has made certain firm-specific investments and possesses firm-specific knowledge (Morris *et al.* 2017), then the worker constitutes a resource. If workers possess firm-specific knowledge, they are not fully substitutable though workers available in external labor markets. With the exception of a few types of workers (e.g. the CEO), workers are a non-scale free resource — their use in one activity prevents their use in another.

How do multibusiness firms decide how to optimally allocate this key resource, workers, across their different business units? Both the theoretical and empirical literature on this specific question is scarce. Existing studies have tended to focus on the internal-to-external transitions of workers (e.g. see literature reviewed in Mawdsley and Somaya (2016)) or internal labor markets as a means of vertically transitioning workers though a firm's hierarchy, i.e. "career ladders" (Doeringer and Piore 1971).

We develop a simple model of the decision to fill labor needs in a business unit by redeploying workers internally instead of hiring them in the external labor market. The model incorporates the assumption that (at least some) workers have firm-specific human capital, i.e. indeed constitute a resource. The model also incorporates key features of the theory of resource redeployment (Helfat and Eisenhardt 2004, Sakhartov and Folta 2014, 2015), such as adjustment costs (in particular, industry relatedness) and industry-level inducements, tailoring them to the specific context of internal labor markets. Finally, we incorporate the existence of external market frictions, in particular cost and rigidities associated with hiring and firing of workers (Lafontaine and Sivadasan 2009, Belenzon and Tsolmon 2016). By explicitly modeling the choice of internal redeployment alongside the alternative of external market resource acquisition, we are precise about the conditions under which internal market transactions are preferable, which the literature has tended to not specifically address.

3.2.1 A Simple Model of Worker Redeployment

In this section, we propose a simple model to gain insights under what conditions a multibusiness firm will staff a labor need in a focal business unit via internal redeployment versus the external labor market.

The firm's objective in any period is to maximize the sum of profits across its business units. We assume, for simplicity, that the firm operates two business units, one in industry j and the other in k, such that $\pi_f = \pi_j + \pi_k$. We assume that demand in each business unit is exogenously given by $d_j = \bar{D}_j / N_j$ where $\bar{D}_j$ is industry demand and N_j is the number of firms in industry j, and that prices are perfectly competitive with $p = 1$. Each business unit j requires one worker of a particular type o (think of type as an occupation, e.g. a welder) who can produce any quantity of output at a constant marginal cost. Labor is the only input into production and the constant marginal cost of a unit of output is $MC_j = w/\tilde{L}_{ij}$ where w is the wage and $\tilde{L}_{ij}$ is the labor productivity of worker i in business unit j. A business unit employing worker i, thus has variable profits of: $\pi_j = d_j(1 - \frac{w}{\tilde{L}_{ij}})$.[4]

The labor productivity of a worker of type o in business unit j is a function of two terms: 1) the worker's general skills (e.g. general expertise, education, experience) and 2) the worker's firm-specific human capital (e.g. tacit knowledge, internal social networks), which we model as: $\tilde{L}_{ij} = h_i s_{if}$. Letting I represent an internal and X an external hire, we assume that $s^I > 1$ and $s^X = 1$, meaning only internal workers have productivity-enhancing

[4]Because we are focused on within-firm, rather than between-firm dynamics, in order to keep the framework simple, we assume perfect competition on the demand side with cost differences on the supply side, without specifying a general equilibrium model of competition between firms where firms price at marginal cost. However, the model conclusions do not hinge on these simplifying assumptions. The key aspects that are required for our analysis are that firms' demand is determined to some extent by external industry conditions and that firms have heterogeneous marginal costs with lower marginal costs mapping into higher profits. Specifying constant electricity of substitution demand with monopolistic competition among firms would also deliver the conditions needed for the analysis.

firm-specific human capital. Thus, we have: $\tilde{L^I}_{ij} = h_i s_{if}$ and $\tilde{L^X}_{ij} = h_i$ with $s_{if} \geq 1$. In what follows, for brevity, we denote π^I_j as the profits generated in a business unit when it employs an internal hire and π^X_j the profits generated by employing an external hire.

A firm looking to hire a unit of labor in business unit j faces two options to fill the position: hiring externally or redeploying the worker internally from k. Hiring for j in the external labor market incurs a one-time hiring cost HC_j (e.g. the costs of time, search) while redeployment from k to j incurs a transfer cost TC_{kj}. Brazilian labor law, for example, guarantees certain rights to employees in the case of internal company transfer to a different address, among them the payment of relocation expenses. If the firm has no slack in k (an assumption to be relaxed later), redeploying the worker also implies replacing her with an external hire in the origin unit k, i.e. incurring a hiring cost in k, HC_k.

Finally, whether a worker of type o is available within the firm is a function of the labor similarity of industries j and k, which we denote by γ_{jk}. For example, if j requires a welder, than γ_{jk} denotes the likelihood that a welder is indeed available in business unit k (e.g. if k is the firm's marketing unit, this probability will be low). This probability will range from zero to one, i.e. $0 \leq \gamma_{jk} \leq 1$.[5] Conditional on a worker of type o being available within the firm (a probability that's increasing in γ_{ij}), then the firm will choose the optimal hiring institution, H^*. The choice between hiring internally (H^I) and hiring externally (H^X) is:

$$H^* = \begin{cases} H^I & \text{if } \pi^I_j + \pi^X_k - TC_{kj} - HC_k \geq \pi^X_j + \pi^I_k - HC_j \\ H^X & \text{otherwise} \end{cases} \tag{3.1}$$

This inequality can also be expressed as, hire internally if:

$$(\pi^I_j - \pi^X_j) + (\pi^X_k - \pi^I_k) \geq TC_{kj} - (HC_j - HC_k) \tag{3.2}$$

Equation 3.2 shows that an internal redeployment will take place if the incremental benefits that an internal worker generates over an external worker in destination unit j

[5]Note that we assume that external labor markets are thick enough that a firm can always find the worker of the needed type.

(term 1) net of the benefits foregone by replacing an internal worker with external one in the origin unit k (term 2) exceed the transfer cost net of any difference in the cost of external hiring in j relative to k.

In the absence of prohibitive hiring costs in j, for a redeployment incentive to exist it has to be the case that the first term of the equation is positive, $\pi_j^I - \pi_j^X > 0$: internal workers have an advantage over external ones in j. Omitting the worker subscript i, this implies: $d_j(1 - \frac{w}{hs_f}) > d_j(1 - \frac{w}{h})$. Assuming for now that internal and external workers with the same general skills are paid the same wage, $w^I = w^X$, this is true when $s_f > 1$, firm-specific human capital advantages exist. Therefore, *in the absence of slack and with zero differences in hiring frictions across locations, workers possessing firm-specific human capital is a necessary condition for redeployment to take place.*

At the same time, note that an internal worker with high levels of firm-specific human capital will also be more valuable in the origin.[6] Indeed, in order for the transfer to be profitable, it has to be the case that the gains of the internal worker in j exceed the opportunity cost of the internal worker at origin k. This condition is met when:

$$d_j(1 - \frac{w}{hs_f}) - d_j(1 - \frac{w}{h}) \geq d_k(1 - \frac{w}{hs_f}) - d_k(1 - \frac{w}{h}) \tag{3.3}$$

which is true when $d_j \geq d_k$—i.e. demand conditions in the destination are weakly better than in the origin.

Thus a *positive difference in demand conditions between the destination and origin is a second necessary condition* for worker redeployment to occur. These observations allow us to formalize our first set of hypotheses:

Hypothesis 1: All else equal, a worker will more likely be redeployed the higher their level of firm-specific human capital. (*firm-specific human capital hypothesis*)

Hypothesis 2: Redeployments will be higher, the more positive the industry conditions of the destination relative to the origin. (*inducement hypothesis*)

[6]Note that we assume that an internal worker is always at least as profitable as an external worker in the origin business unit, i.e. $s_{if} \geq 1$ in the origin.

Hypothesis 3: Redeployments will be higher, the greater the labor relatedness of two industries. (*relatedness hypothesis*)

Revisiting equation 3.2, even if firm-specific knowledge and positive industry differentials exist, the relative benefits of the internal worker in j have to be sufficiently large to compensate for transfer costs. In the case of worker redeployment, transfer costs will include things like reimbursements that a firm has to pay a worker for the cost of moving, as well as any incremental incentives (e.g. bonuses) that the firm will pay to encourage the worker to relocate. In general, geographic distance between plants is likely to imply greater transfer costs, and thus all else equal, fewer redeployments. Finally, per equation 3.2, a differential in the external market hiring costs can also create an incentive for redeployment. In particular, destinations where the external market hiring costs are high, are likely to see more redeployments from within the firm. Seen in a different light, the incremental advantages of an internal workers at the destination can be lower for destinations where the external labor market hiring costs are high. We formalize these observations:

Hypothesis 4: Redeployments will be will be higher the lower the geographic distance between the origin and destination business units. (*distance hypothesis*)

Hypothesis 5: Redeployments will be higher the more unfavorable local labor market conditions of the destination relative to the origin business unit. (*external labor market frictions*)

We next consider a situation where the firm has slack, defined as a need of firing a worker in the origin business unit.[7] This could be due to learning-by-doing, or because the firm is shutting down the origin business unit, for example, due to unfavorable industry conditions.[8] When a firm has a hiring need in j and a slack worker in k, it will redeploy the

[7]We assume that workers are indivisible and use the term "slack" to denote that the worker is superfluous, rather than that she has some excess capacity — i.e. we do not model the possibility of synergies whereby a worker is active in more than one establishment at the same time.

[8]While, to keep the framework simple, we have not modeled any fixed costs of operating, one can imagine the existence of fixed overhead costs in each period, which can lead a firm to decide to shut down when demand is low.

worker internally if:

$$\pi_j^I - TC_{kj} \geq \pi_j^X - HC_j - FC_k \tag{3.4}$$

which can be rewritten as:

$$(\pi_j^I - \pi_j^X) \geq TC_{kj} - (HC_j + FC_k) \tag{3.5}$$

Comparing equations 3.2 and 3.5, we see that *the threshold for redeployment is always lower when the firm has slack*. Note also from equation 3.5, that in some scenarios, the sum of hiring and firing costs may exceed the transfer cost and redeployment may occur even when the right-hand side of equation 3.5 is negative — i.e. when internal workers *do not* have a productivity advantage over external workers in the destination unit. Therefore, redeployments motivated by plant closures may lead to some "inefficient hires" from the perspective of the receiving business unit, which would have been able to attract a higher quality human capital in the absence of firing costs in the origin. Note also that, keeping constant the left side of equation 3.5, a situation involving slack and positive firing costs implies potential for higher levels of transfer costs, compared to a situation of no slack. These observations lead to:

Hypothesis 6a: All else equal, a worker will be more likely to be redeployed if their business unit is exiting. (*slack hypothesis*)

Hypothesis 6b: Redeployments occurring when a business unit is exiting will occur at higher geographic distances, on average, than redeployments occurring when a business unit is not exiting. (*slack-distance trade-off hypothesis*)

Finally, note that thus far, we have assumed that any productivity advantages that internal workers generate due to their firm-specific human capital accrue to the firm via higher profits. In reality, is it likely that firms and workers share the rents generated via some form of Nash bargaining. Therefore, due to their firm-specific human capital, internal hires may earn higher wages than external hires with the same level of general skills:

$w^I > w^X | \, h^I = h^X$.[9] Combining the possibility of excess returns to firm-specific capital with the insights reflected in the prior hypotheses we propose:

Hypothesis 7: A redeployed worker is likely to earn higher wages compared to a worker in the same position and comparable general skills hired externally. The wage premium to the internal worker will be increasing in the worker's level of firm-specific human capital. *(wage advantage hypothesis)*

Overall, this simple model provides several insights into a dual role of internal labor markets. We see that there are two distinct types on inducements to redeployment activity: one, the desire to transfer the "best" workers to their most productive uses, for example in response in differences industry prospects and two, the desire to reduce adjustment costs given slack in an existing business units. These two types of inducements have different implications for which workers are transferred, the wage earnings of the transferred workers, and the productivity advantage of internal versus external hires for the firm. Although redeployments in the absence of slack incentivize the transfer of the workers with the highest levels of firm-specific human capital, redeployments involving slack will be associated with relatively lower levels of human capital and may even result in some "inefficient redeployments" from the perspective of the destination unit. Thus, we see that internal labor markets can play both the function of allocating firm-specific knowledge to its most valuable uses as well as providing an adjustment mechanisms to weather negative shocks, with better-performing different business units absorbing slack generated in business units that under-perform or experience a negative external shock.

3.3 Data and Empirical Strategy

Our primary data source is the Relação Anual de Informações Sociais (RAIS), a mandatory, annual census of all formal-sector employers and their employees in Brazil. These data

[9] Although it's also possible that internal hires may be willing to accept a lower wage than external hires, which would also allow for instances of redeployment driven by this internal wage gap, such cases should not be part of an equilibrium as worker could always quit and receive the (higher) market wage for their skills.

are collected by the Ministry of Labor to support various social insurance programs and contain detailed information on the wages, occupation, and demographics of workers along with the industry and location of employers. Importantly for our purposes, RAIS is an establishment-level census with unique identifiers for each worker, establishment, and firm. We can therefore link workers to firms and observe them moving both between and within firms over time.

We take a ten percent random sample from the population of firms in RAIS that operated establishments in multiple industries between 2004 and 2014. This results in an initial sample of 31,428 establishments in 8,535 unique firms. The average (median) multibusiness firm has 3.6 (2) establishments in 2.1 (2) industries.

To identify instances of worker redeployment, we start with observations from the first and last month that each worker was employed during the year and take the worker's highest paying job within each establishment-month pair. We code redeployment as a worker switching establishments either between the first and last month of employment within a year or from one year to the following calendar year. When analyzing redeployments, we exclude the first and last years of our sample because we cannot observe redeployments occurring between years for the initial and final sample year. This procedure identifies 573,259 worker redeployments for 455,514 unique workers in 7,605 firms over the nine-year period from 2005 to 2013. The final sample of redeployments contains fewer firms than the initial random sample due to the exclusion of the first and last sample years.

Following recent literature (Sakhartov and Folta 2014), our main measure of industry resource relatedness is built from the similarity of industries' occupational requirements. Using data from the year 2000 from RAIS — five years before our sample period — we calculate each of 2,331 different occupations' share of total employment for each industry and then calculate labor relatedness between any two industries as one minus the Euclidean distance of their occupation shares. We then normalize this variable across all the industry pairs to have mean zero and variance equal to one.

To measure the industry opportunities that may act as an inducement for redeploying workers to activities with high returns, we use the two-year growth rate in total industry employment. This measure assumes that greater employment growth within an industry over time is indicative of better prospects for firms. This variable is calculated from RAIS as the total number of unique workers with jobs within a given industry and year across all firms in Brazil.

Tables 3.1, 3.2, and 3.3 show summary statistics for workers, establishments, and destination-origin establishment pairs respectively.

We conduct two types of analyses, one at the level of individual workers and the other at the level of business unit dyads. The first set of worker-level models take the form:

$$redep_{ikt} = \beta + \beta_s s_{it} + \sum_l \beta_l h_{it} + \beta_z z_{kt} + \eta_k + \tau_t + \epsilon_{ikt} \tag{3.6}$$

where the dependent variable takes the value one if a worker was redeployed in the year and zero otherwise. s_{it} is a proxy for a worker's level of firm-specific human capital in year t, h_{it} control for the worker's general level of human capital (education, age, age squared, gender), z_{kt} is establishment size, η and τ are establishment and year fixed effects, and ϵ_{ikt} is a randomly-distributed error term. We perform the analysis using a linear probability model.

The focus of the model is the coefficient on the firm-specific capital proxy, β_s, which conditional of the worker's general skills, estimates the effect of a worker's level of firm-specific human capital on their probability of redeployment. Note that, due to the rich set of fixed effects, the comparison is among workers with different levels of firm experience working in the *same* establishment in the *same* year. We employ two proxies of a worker's level of firm-specific human capital. The first is the worker's years of work experience with the firm. Our second proxy is the worker's position in the occupational hierarchy, i.e. whether the worker's occupation falls in the director/manager, professional, technical and administrative personnel, or production and service worker category. Our prior is that workers higher in the occupational hierarchy are likely to posses rarer and more valuable firm-specific human capital.

Table 3.1: *Worker Summary Statistics*

Variable	mean	sd	p5	p25	p50	p75	p95
Redeployment	0.04	0.18	0	0	0	0	0
Log wage	0.77	0.72	0.01	0.28	0.60	1.09	2.23
Firm experience	2.85	4.16	0	0	1	4	11
Age	31.48	9.77	19	24	29	37	51
Female	0.37	0.48	0	0	0	1	1
Occupation groups							
Managers	0.03	0.18	0	0	0	0	0
Professionals	0.04	0.20	0	0	0	0	0
Technicians & Admin	0.27	0.45	0	0	0	1	1
Service & Production	0.65	0.48	0	0	1	1	1
Education groups							
Below high school	0.40	0.49	0	0	0	1	1
High school	0.51	0.50	0	0	1	1	1
Higher education	0.09	0.29	0	0	0	0	1

Note: Redeployment is a dummy variable for worker redeployment in a given year. Firm experience and age are measured in years.

Table 3.2: *Establishment Summary Statistics*

Variable	mean	sd	p5	p25	p50	p75	p95
Employees	66.4	323.1	1	4	11	33	249
New hires	26.7	151.1	0	1	4	13	96
Separations	23.0	120.9	0	1	3	12	85
Closing year	0.06	0.3	0	0	0	0	1

Note: Separations refers to workers leaving the establishment. Closing year refers to the last year an establishment operates with employees in RAIS.

Table 3.3: *Destination-Origin Dyad Summary Statistics*

Variable	mean	sd	p5	p25	p50	p75	p95
Industry similarity	1.82	1.20	-0.57	1.41	2.44	2.44	2.44
Distance (km)	824	816	18	203	545	1,145	2,683
Difference in growth	0.00	0.12	-0.11	0.00	0.00	0.00	0.11
Closing origin	0.06	0.25	0	0	0	0	1

Note: Difference in growth is calculated as the two-year employment growth rate in the industry of the destination establishment less the same growth rate in the industry of the origin establishment. Closing origin refers to the last year that an origin operates with employees in RAIS.

Based on *H1*, the *specific human capital hypothesis*, we expect β_s to be positive — i.e. workers with higher levels of firm-specific human capital will have a higher likelihood of being redeployed. We also use a slightly modified version of this model to test *H6a* (the *slack hypothesis*) by adding to the model specified in equation 3.6 an indicator variable for whether the worker's current establishment exits in that year. Based on the hypothesis, we expect the coefficient on the exit indicator to be positive and significant.

Our second worker-level model takes workers' contractual wage as the dependent variable. Our tests regarding the wage advantages of redeployed workers over workers hired in the external labor market take the following form:

$$\ln w_{iojt} = \beta + \beta_i redep_{it} + \beta_s s_{it} + \sum_l \beta_l h_{it} + \theta_{ojt} + \epsilon_{iojt} \tag{3.7}$$

The sample of observations for this model are all new employees joining business unit j at time t, which are sourced from either the internal or the external labor market. $redep_{it}$ is an indicator variable taking the value one if the worker was redeployed internally and zero if hired externally. By including a fixed effect for each occupation-establishment-year combination and the full set of worker controls, we are effectively comparing the wage of an internal and an external hire entering the same occupation, in the same establishment, in the same year with the same observable characteristics. Per *H7* (the *wage advantage hypothesis*), we expect β_i to be positive. To further test whether it is the worker's firm-specific human capital that is driving the wage premium, we introduce $\beta_s s_{it}$. The hypothesis is that the wage premium of workers with little firm experience will be small while β_s will be positive.

The final set of models are estimated at the level of business unit dyads. We construct the set of all possible origin- and destination- business units (dyads) within each firm and measure the total amount of redeployments between them in each year (note that each dyad is directional, thus $a \to b$ is not equal to $b \to a$). We estimate the following model:

$$redeps_{k\to j,ft} = \beta + \beta_1 \triangle d_{jkt} + \beta_2 \gamma_{jk} + \beta_3 geo_{jkt} + \beta_4 z_{jt} + \beta_5 z_{kt} + \xi_f + \tau_t + \epsilon_{kjft} \tag{3.8}$$

where $redeps_{k\to j,ft}$ is the number of workers redeployed from origin k to destination j within

Table 3.4: *Percentage of Workers Redeployed by Establishment*

Condition	mean	sd	p50	p75	p90	p95	p99
Incoming redeployments as percentage of							
Employees	5.5	15.2	0.0	2.5	15.4	33.3	91.3
New Hires	12.1	24.0	0.0	12.5	50.0	71.4	100
New Plants	22.5	32.8	0.0	40.0	81.8	100	100
Outgoing redeployments as percentage of							
Employees	4.8	13.8	0.0	2.2	13.6	25.0	83.3
Separations	11.8	23.9	0.0	12.5	44.4	68.8	100
Closing Plants	21.8	35.2	0.0	33.3	100	100	100

NOTE: Numbers represent redeployments as a percentage of employees in each category. Incoming redeployments are workers joining an establishment from another establishment owned by the same firm; outgoing redeployments are workers leaving an establishment for another establishment of the same firm. Numbers for hires and separations (i.e. workers leaving the establishment) are conditional on anyone being hired or exiting.

firm f in year t, $\triangle d_{jkt}$ is a measure of the difference in prospects of j and k's industries, γ_{jk} is the industry relatedness, geo_{jkt} is geographic distance between the plants and the zs are establishment controls. The model also includes firm- and year fixed effects. To test *H2* (the *inducement hypothesis*), we focus on the sign of β_1 which is expected to be positive. We test *H3* (the *relatedness hypotheses*) by evaluating the sign on β_2 which is expected to be positive. *H4* (the *distance hypothesis*) predicts that the sign of β_3 is negative. In other versions of this model, we also introduce an indicator for whether the origin business unit exits at t and its interaction with the main effects, to test *H6b*, whether redeployments occur at larger distances when the origin plant is exiting.

3.4 Results

Worker redeployment is pervasive. Table 3.4 shows that for multibusiness firms in Brazil, on average 12.1 percent of all new hires come from other establishments of the firm. Among workers leaving an establishment, 11.8 percent move to jobs within the same firm. This percentage is even higher when firms close an establishment; 21.8 percent of workers in establishments that are ceasing operations move to new positions within the same firm.

Employees higher in the organizational hierarchy and employees in professional roles are

Table 3.5: *Percentage of Workers Redeployed by Occupation*

Occupation	All Years	Closing Plants	New Plants
Managers	8.2	27.7	41.6
Professionals	6.4	30.2	34.1
Technicians & Admin	5.0	23.4	25.4
Service & Production	4.4	21.3	20.5
Total	5.6	23.7	27.0

NOTE: Numbers represent redeployments as a percentage of employees in each category. *Closing Plants* refers to establishments in their final year of operation and numbers represent the percentage of employees who move to another establishment of the same firm. *New Plants* refers to the first year of a new establishment and numbers represent the percentage of employees hired from other establishments of the same firm.

more likely than others to be redeployed between establishments of the same firm. Table 3.5 shows that on average, 8.2 percent of an establishment's managers are redeployed in any year. This is nearly double the 4.4 percent of service and production workers that are redeployed. The gap in redeployment between managers and others narrows when establishments close. On average, closing establishments redeploy 28 percent of their managers, 30 percent of their professionals, and roughly 22 percent of other workers. Managers' increased likelihood of redeployment is suggestive evidence for the hypothesis that firms use redeployment as a tool for reallocating valuable human capital resources.

Consistent with the hypothesis that redeployment is increasing in firm-specific human capital, Table 3.6 shows an additional year of experience working in a firm increases the probability of redeployment by roughly 0.1 percentage points. About 3.5 percent of all workers are redeployed in any given year, so a 0.1 percentage point increase represents a 2.9 percent increase in the probability of redeployment. Model 2 of Table 3.6 provides further support for the hypothesis. Workers who are more likely to have valuable firm-specific human capital — managers and professionals — are much more likely to be redeployed. Controlling for worker characteristics and establishment and year fixed effects, managers are redeployed at a rate that is nearly five percentage points higher than the probability of redeployment for service and production workers.

Closing an establishment dramatically increases the probability that workers will be redeployed. As hypothesized, Table 3.6 shows that closing an establishment is associated

Table 3.6: *Redeployment of Workers*

	(1)	(2)	(3)	(4)	(5)
Firm experience	0.0009***		0.0007***	0.0008***	
	(0.0003)		(0.0002)	(0.0003)	
Managers		0.0481***	0.0466***		0.0466***
		(0.0118)	(0.0113)		(0.0121)
Professionals		0.0091**	0.0086*		0.0076*
		(0.0045)	(0.0044)		(0.0046)
Technicians & Admin		0.0069***	0.0064***		0.0059***
		(0.0012)	(0.0012)		(0.0012)
Closing				0.1353***	0.1315***
				(0.0321)	(0.0309)
Closing ×					
Firm experience				0.0045**	
				(0.0020)	
Managers					0.0815***
					(0.0248)
Professionals					0.1137***
					(0.0313)
Technicians & Admin					0.0609***
					(0.0226)
High school	0.0047***	0.0029***	0.0031***	0.0047***	0.0028***
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Higher Ed	0.0161***	0.0071	0.0075*	0.0163***	0.0072
	(0.0042)	(0.0045)	(0.0043)	(0.0042)	(0.0045)
Age	0.0013***	0.0014***	0.0012***	0.0013***	0.0013***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Age squared	-0.0000***	-0.0000***	-0.0000***	-0.0000***	-0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Female	-0.0031***	-0.0031***	-0.0029***	-0.0032***	-0.0031***
	(0.0011)	(0.0010)	(0.0009)	(0.0011)	(0.0009)
Log employment	0.0141***	0.0139***	0.0144***	0.0195***	0.0193***
	(0.0049)	(0.0049)	(0.0049)	(0.0051)	(0.0051)
Firms	7,474	7,474	7,474	7,474	7,474
Observations	6,179,313	6,179,313	6,179,313	6,179,313	6,179,313

NOTE: Standard errors in parentheses are clustered by firm. All models include establishment and year fixed effects. The excluded category for the education dummies is "Less than high school"; the excluded category for the occupation categories is "Service & Production" workers.
* $p < .10$, ** $p < .05$, *** $p < .01$

with a 13.5 percentage point increase in the probability of redeployment; this is a more than 300 percent increase in the probability of redeployment. Models 4 and 5 further show that the impact of closing an establishment is even greater for employees with more firm-specific work experience and those higher in the organizational hierarchy — e.g. managers. Intuitively, an establishment closing represents an opportunity for the loss of rents from firm-specific human capital. In these cases, redeployment allows the firm to keep a worker within its boundaries and therefore maintain any benefits of firm-specific human capital.

Workers who are redeployed (i.e. internal hires) earn a wage premium over workers who are hired externally, and the premium is increasing in the firm-specific experience of redeployed workers. Models 1–4 of Table 3.7 compare the contractual wage of internal and external hires, controlling for an establishment-occupation-year fixed effect.[10] This model compares workers hired for the same occupation, in the same establishment, and in the same year. On average, redeployed workers — i.e. those hired internally from another unit of the same firm — earn about nine percent more than workers hired from other firms (model 1). Model 2 shows that this wage premium is increasing in the firm-specific work experience of redeployed workers. Specifically, a worker redeployed in their first year with zero years of firm-specific experience earns an average wage premium of 3.5 percent over external hires. This premium increases by roughly 2.4 percent for each year of experience working within the firm. Model 3 suggests that workers hired from closing establishments within the same firm, however, earn much lower wage premiums over external hires than workers moving from establishments that are not shutting down. The comparison between workers moving from closing versus continuing establishments should be interpreted cautiously; the use of establishment-occupation-year fixed effects in the model means that this comparison depends on establishments hiring internal workers from both closing and ongoing establishments to perform the same job in the same year.[11]

[10]The RAIS data provide information on several different measures of worker compensation, including the contractual wage and the actual amounts paid out to workers in any given year. While the two are highly correlated, for our analysis we use the contractual wage in order to not capture any non-recurring payments that may otherwise be included in that year's wage for redeployed workers, such as relocation bonuses.

[11]There are 1,690 establishment-occupation-year cells with this variation (out of approximately 1.8 million

Table 3.7: *Wages of Redeployed Workers*

	(1)	(2)	(3)	(4)	(5)
Redeploy	0.093***	0.032***	0.036***	0.095***	0.012
	(0.011)	(0.009)	(0.010)	(0.015)	(0.011)
Firm exp.		0.024***	0.024***		0.016***
		(0.004)	(0.004)		(0.001)
Closing origin			-0.041***		
			(0.014)		
Redeploy ×					
Firm exp.					0.006**
					(0.003)
Managers				0.014	
				(0.021)	
Professionals				-0.011	
				(0.025)	
Technicians and Admin				-0.008	
				(0.017)	
High school	0.016***	0.017***	0.017***	0.016***	0.041***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Higher Ed	0.174***	0.177***	0.177***	0.174***	0.252***
	(0.013)	(0.012)	(0.012)	(0.013)	(0.013)
Age	0.004***	0.004***	0.004***	0.004***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age squared	-0.000*	-0.000**	-0.000**	-0.000*	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.021***	-0.021***	-0.021***	-0.021***	-0.047***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Firms	5,991	5,991	5,991	5,991	6,749
Observations	1,833,356	1,833,356	1,833,356	1,833,356	3,421,007

Note: All models include establishment-occupation-year fixed effects so that comparisons between redeployed workers and other workers are within establishment-occupation-year. The excluded category for the education dummies is "Less than high school"; the excluded category for the occupation categories is "Service & Production" workers. Standard errors in parentheses are clustered by firm.
* $p < .10$, ** $p < .05$, *** $p < .01$

Table 3.8: *Adjustment Costs and Redeployment*

	(1)	(2)	(3)	(4)
Diff. in growth	0.017			0.018
	(0.017)			(0.017)
Industry similarity		0.029***		0.037***
		(0.010)		(0.011)
Log distance			-0.125***	-0.125***
			(0.013)	(0.013)
Dest. log employment	0.074***	0.075***	0.082***	0.083***
	(0.011)	(0.011)	(0.010)	(0.010)
Origin log employment	0.138***	0.138***	0.149***	0.150***
	(0.024)	(0.024)	(0.023)	(0.023)
Firms	2,286	2,286	2,282	2,282
Observations	59,266	59,264	59,219	59,217

NOTE: Observations are establishment dyads (i.e. a destination and origin establishment pair) with positive redeployment. The dependent variable is the natural logarithm of redeployments. All models include firm, destination industry, origin industry, and year fixed effects. Standard errors in parentheses are clustered by firm.
* $p < .10$, ** $p < .05$, *** $p < .01$

One possible explanation for the large wage premium of internal over external hires is unobservable differences in skill between those who are hired internally through redeployment and workers hired through external labor markets. To address this possibility, Model 5 of Table 3.7 compares the wages of redeployed workers and other workers at the destination who were not hired that year. In other words, unlike models 1–4, model 5 compares redeployed workers to their peers who were not redeployed. These results show no statistically significant wage premium for redeployment in the absence of firm-specific experience. This suggests that redeployed workers in fact resemble other workers in their destination establishment. The interaction of redeployment with firm-specific experience, however, indicates that redeployed workers may earn an additional premium for firm-specific work experience. One additional year of firm-specific experience increases the wages of workers who are not redeployed by roughly 1.5 percent, versus a 2.1 percent increase for those who are redeployed.

Firms redeploy workers more intensively between establishments in related industries

total observations.

Table 3.9: *Adjustment Costs and Redeployment When Closing*

	(1)	(2)	(3)
Closing origin	0.508***	0.421***	0.728***
	(0.081)	(0.053)	(0.106)
Closing origin ×			
Industry similarity		0.050*	
		(0.030)	
Log distance			-0.060***
			(0.018)
Industry similarity	0.036***	0.032***	0.036***
	(0.010)	(0.011)	(0.010)
Log distance	-0.124***	-0.124***	-0.121***
	(0.013)	(0.013)	(0.013)
Diff. in growth	0.014	0.015	0.014
	(0.018)	(0.018)	(0.018)
Dest. log employment	0.080***	0.081***	0.081***
	(0.010)	(0.010)	(0.010)
Origin log employment	0.174***	0.173***	0.174***
	(0.020)	(0.020)	(0.020)
Firms	2,282	2,282	2,282
Observations	59,217	59,217	59,217

Note: Observations are establishment dyads (i.e. a destination and origin establishment pair) with positive redeployment. The dependent variable is the natural logarithm of redeployments. All models include firm, destination industry, origin industry, and year fixed effects. Standard errors in parentheses are clustered by firm.
* $p < .10$, ** $p < .05$, *** $p < .01$

and establishments that are geographically closer to each other. Table 3.8 shows that a one standard deviation increase in industry similarity is associated with a three percent increase in redeployments. Models 3 and 4 show that a one percent increase in distance between establishments is associated with 0.13 percent fewer redeployments. Together, these results support the relatedness and distance hypotheses (i.e. hypotheses 3 and 4). The results for the inducement hypothesis that favorable industry conditions in a destination establishment relative to an origin establishment are equivocal. The coefficients in models 1 and 4 of Table 3.8 have the expected sign, but are not statistically significant.

Consistent with the results of Table 3.6 showing an increased likelihood of redeployment when closing an establishment, models 1–3 of Table 3.9 show that the intensity of redeployment is also greater when closing an establishment even after controlling for industry similarity, geographic distance, differences in industry growth, and establishment size. Specifically, a closing origin establishment is associated with a roughly 50 percent increase in the number of worker redeployments (see model 1). Models 2–3, however, fail to support the hypotheses that redeployments occurring when an establishment closes (i.e. under slack conditions) will be less sensitive to industry relatedness and geographic distance. The coefficients on the interactions between an establishment closing and industry similarity or geographic distance do not have the expected sign; the positive and negative coefficients on these interactions respectively suggest that firms may be *more* sensitive to industry relatedness and geographic distance when redeploying workers from a closing establishment. Some caution is warranted, however, when interpreting these results because workers redeployed when establishments close may be unlike workers redeployed during normal business conditions. Specifically, such workers may be less adaptable to new industries or more sensitive to moving large geographic distances.

3.5 Conclusion

This study has explored the extent and drivers of internal labor market activity in multibusiness firms in the context of Brazil. We have presented a simple framework where two distinct forces can give rise to an incentive to redeploy workers: external labor market frictions (hiring and firing costs) and workers' possession of valuable, firm-specific knowledge.

We find that Brazilian multibusiness firms source a meaningful share of their workers from within the firm. In an average year, the typical establishment sources 12.1 percent of new hires internally. Studying what predicts whether a worker is redeployed, we find evidence that both workers higher in the occupational hierarchy and workers with more firm-specific experience are more likely to be redeployed. Managers, in particular, are redeployed more than twice as often as the average worker.

We also study the wages of workers hired into a position through internal redeployment versus the external labor market. Comparing two workers hired into the same narrow occupation, in the same establishment, in the same year, with otherwise similar characteristics, we find that redeployed workers earn a nine percent wage premium over those hired externally. On the other hand, redeployed workers do not earn a substantial premium over otherwise comparable workers at the destination, suggesting that these results are not driven by selection on unobservable worker quality (redeployed workers being of better quality relative to other internal workers). The wage premium is consistent with firm-specific experience rather than worker's personal motivations or external hiring frictions driving redeployment.

Our paper contributes to existing theory of resource redeployment, which has theorized but rarely observed actual redeployment. Our results show that redeployment is pervasive in the context of internal labor markets in multibusiness firms, and most often does not involve the simultaneous exit of the origin business unit. Furthermore, our paper contributes to the broader literature on internal labor markets. Compared to the existing focus on vertical labor markets and horizontal labor markets as a response to external labor market frictions, the results of our paper support the view that internal labor markets also serve as conduits of firm-specific knowledge inside the firm.

Our results also provide directions for future research. One feature that we have observed in the data is that redeployments are especially high when firms first open new establishments. Tables 3.4 and 3.5 show that in such cases, 22.5 percent of all initial workers and 42 percent of all managers of the new plants are sourced from other units of the same firm. Understanding the strategies multibusiness firms use when they engage in "intrapreneurship" — in particular the type and nature of the human resources allocated to new businesses — and whether the option to leverage their internal labor markets provides a competitive advantage over independent startups constitutes an important and interesting area for future research.

References

Agrawal, A., Galasso, A. and Oettl, A. (2017). Roads and Innovation. *Review of Economics and Statistics*, **99** (3), 417–434.

Aitken, B. and Harrison, A. E. (1999). Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela. *American Economic Review*, **89** (3), 605–618.

Alcácer, J. (2006). Location Choices Across the Value Chain: How Activity and Capability Influence Collocation. *Management Science*, **52** (10), 1457–1471.

Alcácer, J. and Delgado, M. (2016). Spatial Organization of Firms and Location Choices Through the Value Chain. *Management Science*, **62** (11), 3213–3234.

— and Zhao, M. (2016). Zooming in: A Practical Manual for Identifying Geographic Clusters. *Strategic Management Journal*, **37** (1), 10–21.

Alfaro, L. and Charlton, A. (2009). Intra-Industry Foreign Direct Investment. *American Economic Review*, **99** (5), 2096–2119.

— and Chen, M. X. (2014). The Global Agglomeration of Multinational Firms. *Journal of International Economics*, **94** (2), 263–276.

— and — (2017). Selection and Market Reallocation: Productivity Gains from Multinational Production. *American Economic Journal: Economic Policy*.

Allen, T. and Arkolakis, C. (2014). Trade and the Topography of the Spatial Economy. *Quarterly Journal of Economics*, **129** (3), 1085–1139.

Almeida, P., Song, J. and Grant, R. M. (2002). Are Firms Superior to Alliances and Markets? An Empirical Test of Cross-Border Knowledge Building. *Organization Science*, **13** (2), 147–161.

Arikan, A. T. and Schilling, M. A. (2011). Structure and Governance in Industrial Districts: Implications for Competitive Advantage. *Journal of Management Studies*, **48** (4), 772–803.

Asturias, J., García-Santana, M. and Ramos, R. (2015). Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India.

Atalay, E., Hortacsu, A., Li, M. and Syverson, C. (2017). How Wide is the Firm Border?

—, Hortaçsu, A. and Syverson, C. (2014). Vertical Integration and Input Flows. *American Economic Review*, **104** (4), 1120–1148.

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, **17** (1), 99–120.

Basile, R., Pittiglio, R. and Reganati, F. (2017). Do Agglomeration Externalities Affect Firm Survival? *Regional Studies*, **51** (4), 548–562.

Baum, J. a. C. and Haveman, H. a. (1997). Love Thy Neighbor? Differentiation and Agglomeration in the Manhattan Hotel Industry, 1898–1990. *Administrative Science Quarterly*, **42** (2), 304–338.

Baum, J. A. C. and Mezias, S. J. (1992). Localized Competition and Organizational Failure in the Manhattan Hotel Industry, 1898–1990. *Administrative Science Quarterly*, **37** (4), 580–604.

Beaudry, C. and Swann, G. M. P. (2009). Firm growth in industrial clusters of the United Kingdom. *Small Business Economics*, **32** (4), 409–424.

Bel, G. and Fageda, X. (2007). Getting there fast: globalization, intercontinental flights and location of headquarters. *Journal of Economic Geography*, **8** (4), 471–495.

Belenzon, S. and Tsolmon, U. (2016). Market Frictions and the Competitive Effects of Internal Labor Markets. *Strategic Management Journal*, **37** (2), 315–334.

Bernard, A. B., Jensen, J. B., Redding, S. J. and Schott, P. K. (2007). Firms in International Trade. *Journal of Economic Perspectives*, **21** (3), 105–130.

Bernard, B. A. B., Redding, S. J. and Schott, P. K. (2010). Multiple-Product Firms and Product Switching. *American Economic Review*, **100** (1), 70–97.

Boeh, K. K. and Beamish, P. W. (2012). Travel time and the liability of distance in foreign direct investment: Location choice and entry mode. *Journal of International Business Studies*, **43** (5), 525–535.

Buenstorf, G. and Klepper, S. (2009). Heritage and Agglomeration: The Akron Tyre Cluster Revisited. *Economic Journal*, **119** (537), 705–733.

Cestone, G., Pica, G., Fumagalliz, C. and Francis Kramarz (2017). Insurance Between Firms: The Role of Internal Labor Markets.

Choudhury, P. (2017). Innovation Outcomes in a Distributed Organization: Intrafirm Mobility and Access to Resources. *Organization Science*, **28** (2), 339–354.

Chung, W. and Alcácer, J. (2002). Knowledge Seeking and Location Choice of Foreign Direct Investment in the United States. *Management Science*, **48** (12), 1534–1554.

— and Kalnins, A. (2001). Agglomeration Effects and Performance: A Test of the Texas Lodging Industry. *Strategic Management Journal*, **22** (10), 969–988.

Combes, P. P., Duranton, G. and Gobillon, L. (2011). The identification of agglomeration economies. *Journal of Economic Geography*, **11** (2), 253–266.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D. and Roux, S. (2012). The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica*, **80** (6), 2543–2594.

Delgado, M., Porter, M. E. and Stern, S. (2014). Clusters, Convergence, and Economic Performance. *Research Policy*, **43** (10), 1785–1799.

—, — and — (2015). Defining clusters of related industries. *Journal of Economic Geography*, **16** (1), 1–38.

Doeringer, P. and Piore, M. (1971). *Internal Labor Markets and Manpower Analysis*. Lexington, MA: Heath Lexington Books.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, **108** (4-5), 899–934.

Duranton, G. and Overman, H. (2005). Testing for Localization Using Micro-Geographic Data. *The Review of Economic Studies*, **72** (4), 1077–1106.

Ellison, G. and Glaeser, E. L. (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, **105** (5), 889.

—, — and Kerr, W. R. (2010). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, **100** (3), 1195–1213.

Faber, B. (2014). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies*, **81** (3), 1046–1070.

Fama, E. F. (1980). Agency Problems and the Theory of the Firm. *Journal of Political Economy*, **88** (2), 288–307.

Fan, J. and Lang, L. (2000). The Measurement of Relatedness: An Application to Corporate Diversification. *The Journal of Business*, **73** (4), 629–660.

Farjoun, M. (1994). Beyond Industry Boundaries: Human Expertise, Diversification and Resource-Related Industry Groups. *Organization Science*, **5** (2), 185–199.

Folta, T. B., Cooper, A. C. and Baik, Y. S. (2006). Geographic cluster size and firm performance. *Journal of Business Venturing*, **21** (2), 217–242.

—, Helfat, C. E. and Karim, S. (2016). Examining resource redeployment in multi-business firms. In *Resource Redeployment and Corporate Strategy*, Emerald Group Publishing Limited, pp. 1–17.

Frenken, K., Cefis, E. and Stam, E. (2014). Industrial Dynamics and Clusters: A Survey. *Regional Studies*, **49** (1), 10–27.

Ghemawat, P. (2007). *Redefining global strategy: Crossing borders in a world where differences still matter*. Cambridge, MA: Harvard Business Press.

— and Thomas, C. (2008). Strategic Interaction Across Countries and Multinational Agglomeration: An Application to the Cement Industry. *Management Science*, **54** (12), 1980–1996.

Gibbons, S., Lyytikäinen, T., Overman, H. and Sanchis-Guarner, R. (2017). New Road Infrastructure: the Effects on Firms. *SERC Discussion paper*.

Giroud, X. (2012). Proximity and Investment: Evidence from Plant-Level Data. *The Quarterly Journal of Economics*, **128** (2), 861–915.

— and Mueller, H. M. (2014). Capital and Labor Reallocation within Firms. *The Journal of Finance*, **40** (4).

Goswami, A. G., Ghani, E. and Kerr, W. R. (2016). Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing. *The Economic Journal*, **126** (591), 317–357.

Grant, R. M. (1996). Toward a Knowledge-Based Theory of the Firm. *Strategic Management Journal*, **17**, 109–122.

Gross, D. P. (2016). The Ties that Bind: Railroad Gauge Standards and Internal Trade in the 19th Century U.S.

Groysberg, B., Lee, L.-E. and Nanda, A. (2008). Can They Take It With Them? The Portability of Star Knowledge Workers' Performance. *Management Science*, **54** (7), 1213–1230.

Haltiwanger, J. C., Lane, J. I. and Spletzer, J. R. (1999). Productivity Differences across Employers: The Roles of Employer Size, Age, and Human. *The American Economic Review*, **89** (2), 94–98.

Hannan, M. T. and Freeman, J. (1977). The Population Ecology of Organizations. *American Journal of Sociology*, **82** (5), 929–964.

Haveman, H. a. and Rider, C. I. (2014). The Spatial Scope of Competition and the Geographical Distribution of Entrepreneurship: Magazine Foundings and the U.S. Post Office. *Sociological Science*, **1**, 111–127.

Helfat, C. E. and Eisenhardt, K. M. (2004). Inter-Temporal Economies of Scope, Organizational Modularity, and the Dynamics of Diversification. *Strategic Management Journal*, **25** (13), 1217–1232.

Henderson, J. V. (2003). Marshall's scale economies. *Journal of Urban Economics*, **53** (1), 1–28.

Henisz, W. J. (2002). The institutional environment for infrastructure investment. *Industrial and Corporate Change*, **11** (2), 355–389.

Hotelling, H. (1929). Stability in Competition. *The Economic Journal*, **39** (153), 41–57.

Kalnins, A. and Chung, W. (2004). Resource-Seeking Agglomeration: A Study of Market Entry in the Lodging Industry. *Strategic Management Journal*, **25** (7), 689–699.

— and Lafontaine, F. (2013). Too Far Away? The Effect of Distance to Headquarters on Business Establishment Performance. *American Economic Journal: Microeconomics*, **5** (3), 157–179.

Karim, S. and Mitchell, W. (2004). Innovating through Acquisition and Internal Development: A Quarter-century of Boudary Evolution at Johnson & Johnson. *Long Range Planning*, **37** (6), 525–547.

Keller, W. and Yeaple, S. (2013). The Gravity of Knowledge. *American Economic Review*, **103** (4), 1414–1444.

Kerr, W. R. and Duranton, G. (2015). The Logic of Agglomeration.

— and Kominers, S. D. (2015). Agglomerative Forces and Cluster Shapes. *Review of Economics and Statistics*, **97** (4), 877–899.

Khanna, T. and Palepu, K. (2000). Is Group Affiliation Profitable in Emerging Markets? An Analysis of Diversified Indian Business Groups. *The Journal of Finance*, **55** (2), 867–891.

— and Palepu., K. G. (2010). *Winning in emerging markets: A road map for strategy and execution.* Harvard Business Press.

Kogut, B. and Zander, U. (1992). Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organization Science*, **3** (3), 383–397.

Lafontaine, F. and Sivadasan, J. (2009). Do Labor Market Rigidities Have Microeconomic Effects? *American Economic Journal: Applied Economics*, **1** (2), 88–127.

Lahiri, N. (2010). Geographic Distribution of R&D Activity: How Does It Affect Innovation Quality? *Academy of Management Journal*, **53** (5), 1194–1209.

Lamont, O. (1997). Cash Flow and Investment: Evidence from Internal Capital Markets. *The Journal of Finance*, **52** (1), 83–109.

Lemelin, A. (1982). Relatedness in the Patterns of Industry Diversification. *Review of Economics and Statistics*, **64** (4), 646–657.

Levinthal, D. A. and Wu, B. (2010). Opportunity Costs and Non-Scale Free Capabilities: Profit Maximization, Corporate Scope, and Profit Margins. *Strategic Management Journal*, **31** (7), 780–801.

Lieberman, M., Lee, G. and Folta, T. (2016). Entry, Exit, and the Potential for Resource Redeployment. *Strategic Management Journal*, **38** (3), 526–544.

Lippman, S. A. and Rumelt, R. P. (1982). Uncertain Imitability: An Analysis of Interfirm Differences in Efficiency under Competition. *The Bell Journal of Economics*, **13** (2), 418–438.

Marshall, A. (1920). *Principles of Economics*. London: MacMillan.

Martin, P., Mayer, T. and Mayneris, F. (2011). Spatial Concentration and Plant-Level Productivity in France. *Journal of Urban Economics*, **69** (2), 182–195.

Martin, X., Salomon, R. and Wu, Z. (2010). The Institutional Determinants of Location Choice: Agglomeration in the Global Semiconductor Industry. *Industrial and Corporate Change*, **19** (6), 1769–1800.

Mawdsley, J. K. and Somaya, D. (2016). Employee Mobility and Organizational Outcomes: An Integrative Conceptual Framework and Research Agenda. *Journal of Management*, **42** (1), 85–113.

McCann, B. T. and Folta, T. B. (2008). Location Matters: Where We Have Been and Where We Might Go in Agglomeration Research. *Journal of Management*, **34** (3), 532–565.

— and — (2011). Performance differentials within geographic clusters. *Journal of Business Venturing*, **26** (1), 104–123.

Melitz, M. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, **71** (6), 1695–1725.

Mian, A. and Sufi, A. (2014). What Explains the 2007–2009 Drop in Employment? *Econometrica*, **82** (6), 2197–2223.

Mingo, S. (2013). The Impact of Acquisitions on the Performance of Existing Organizational Units in the Acquiring Firm: The Case of an Agribusiness Company. *Management Science*, **59** (12), 2687–2701.

Montgomery, C. A. and Hariharan, S. (1991). Diversified expansion by large established firms. *Journal of Economic Behavior and Organization*, **15** (1), 71–89.

Morris, S. S., Alvarez, S. A., Barney, J. B. and Molloy, J. C. (2017). Firm-specific human capital investments as a signal of general value: Revisiting assumptions about human capital and how it is managed. *Strategic Management Journal*, **38** (4), 912–919.

Morten, M. and Oliveira, J. (2016). The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City.

Nachum, L. and Zaheer, S. (2005). The persistence of distance? The impact of technology on MNE motivations for foreign investment. *Strategic Management Journal*, **26** (8), 747–767.

Pe'er, A., Vertinsky, I. and Keil, T. (2016). Growth and survival: The moderating effects of local agglomeration and local market structure. *Strategic Management Journal*, **37**, 541–564.

Penrose, E. (1959). *The Theory of the Growth of the Firm*. Oxford University Press.

Pinkse, B. Y. J., Slade, M. E. and Brett, C. (2002). Spatial Price Competition: A Semiparametric. *Econometrica*, **70** (3), 1111–1153.

Porter, M. E. (1990). *The Competitive Advantage of Nations*. New York: The Free Press.

Ricart, J. E., Enright, M. J., Ghemawat, P., Hart, S. L. and Khanna, T. (2004). New Frontiers in International Stratey. *Journal of International Business Studies*, **35** (3), 175–200.

Rigby, D. L. and Brown, W. M. (2015). Who Benefits from Agglomeration? *Regional Studies*, **49** (1), 28–43.

Rosenthal, S. and Strange, W. (2003). Geography, Industrial Organization, and Agglomeration. *Review of Economics and Statistics*, **85** (2), 377–393.

Rumelt, R. P. (1982). Diversification Strategy and Profitability. *Strategic Management Journal*, **3** (4), 359–369.

Sakhartov, A. V. and Folta, T. (2014). Resource Relatedness, Redeployability, and Firm Value. *Strategic Management Journal*, **35**, 1781–1797.

— and — (2015). Getting Beyond Relatedness as a Driver of Corporate Value. *Strategic Management Journal*, **51** (2), 315–334.

Salop, S. C. (1979). Monopolistic Competition with Outside Goods. *The Bell Journal of Economics*, **10** (1), 141–156.

Schettini, D. and Azzoni, C. (2015). Determinantes Regionais da Productividade Industrial: O Papel da Infraestrutura. In *Produtividade no Brasil*, vol. 2, Instituto de Pesquisa Econômica Aplicada (Ipea) Organizadores, pp. 391–415.

Schoar, A. (2002). Effects of Corporate Diversification on Productivity. *The Journal of Finance*, **57** (6), 2379–2403.

Shaver, M. J. and Flyer, F. (2000). Agglomeration Economies, Firm Heterogeneity, and Foreign Direct Investment in the United States. *Strategic Management Journal*, **21** (12), 1175–1193.

Shin, H.-H. and Stulz, R. M. (1998). Are Internal Capital Markets Efficient? *The Quarterly Journal of Economics*, **113** (2), 531–552.

Silverman, B. S. (1999). Technological Resources and the Direction of Corporate Diversification: Toward an Integration of the Resource-Based View and Transaction Cost Economics. *Management Science*, **45** (8), 1109–1124.

Sorenson, O. and Audia, P. G. (2000). The Social Structure of Entrepreneurial Activity: Geographic Concentration of Footwear Production in the United States 1940-1989. *American Journal of Sociology*, **106** (2), 424–462.

Stein, J. C. (1997). Internal Capital Markets and the Competition for Corporate Resources. *The Journal of Finance*, **52** (1), 111–133.

Strauss-Kahn, V. and Vives, X. (2009). Why and Where do Headquarters Move? *Regional Science and Urban Economics*, **39** (2), 168–186.

Syverson, C. (2004). Market Structure and Productivity: A Concrete Example. *Journal of Political Economy*, **112** (6), 1181–1222.

Tate, G. and Yang, L. (2015). The Bright Side of Corporate Diversification: Evidence from Internal Labor Markets. *Review of Financial Studies*, **28** (8), 2203–2249.

Teece, D. (1982). Towards an Economic Theory of the Multiproduct Firm. *Journal of Economic Behavior & Organization*, **3** (1), 39–62.

Teece, D. J. (1981). The Market for Know-How and the Efficient International Transfer of Technology. *The ANNALS of the American Academy of Political and Social Science*, **458** (1), 81–96.

Vernon Henderson, J. (2003). Marshall's Scale Economies. *Journal of Urban Economics*, **53**, 1–28.

Wang, R. and Shaver, M. (2013). Competition-Driven Repositioning. *Strategic Management Journal*, **35** (11), 1585–1604.

Wernerfelt, B. (1984). A Resource-Based View of the Firm. *Strategic Management Journal*, **5** (2), 171–180.

Zaheer, S. (1995). Overcoming the Liability of Foreignness. *Academy of Management Journal*, **38** (2), 341–363.

Appendix A

A.1 Supplementary Tables for Chapter 1

Table A.1: *Firm Descriptive Statistics in 2007*

	All firms		Locally traded		Nationally traded	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Number of workers	23.307	86.272	18.525	53.039	39.002	169.143
Age (years)	8.412	5.099	8.151	5.034	8.858	5.172
Exporter	0.050	0.218	0.029	0.166	0.088	0.284
Importer	0.048	0.214	0.028	0.165	0.083	0.275
Survive	0.556	0.497	0.557	0.497	0.554	0.497
Relocate	0.037	0.188	0.031	0.173	0.045	0.207
Switch product	0.107	0.309	0.084	0.278	0.110	0.312
N (number of firms)	129,325		53,289		22,726	

Note: Author's calculations based on data from Relação Anual de Informações Sociais (RAIS) and the Secretaria de Comércio Exterior (SECEX).

Table A.2: *Federal Roads by Surface Type in 2007 and 2014*

Surface type	2007 Length (km)	2007 *Percent*	2014 Length (km)	2014 *Percent*
Two-lane highway	7,234	7%	9,276	8%
Being duplicated	1,079	1%	1,680	2%
Paved road	76,420	72%	80,613	74%
Being paved	4,822	5%	4,482	4%
Unpaved (year-round)	12,311	12%	9,090	8%
Being fortified	236	0%	189	0%
Dirt road	3,449	3%	4,180	4%

Note: Author's calculations based on data from the National Department of Transportation Infrastructure (DNIT).

Table A.3: *Federal Roads by Surface Condition in 2007 and 2014*

Condition	2007 Length (km)	2007 *Percent*	2014 Length (km)	2014 *Percent*
Excellent	3,860	4%	7,040	8%
Good	6,748	7%	26,587	29%
Regular	65,638	71%	39,597	43%
Poor	12,245	13%	15,658	17%
Very poor	3,837	4%	3,445	4%

Note: Author's calculations based on data from the Brazilian National Transport Confederation (CNT).

Table A.4: *Assumed Velocities (km/hr) Given Road Surface and Condition*

		Two-lane highway	Being duplicated	Paved road	Being paved	Unpaved (year-round)	Being fortified	Dirt road
	Factor	1.20	1.10	1.00	0.80	0.60	0.50	0.40
Excellent	1.10	106	97	88	70	53	44	35
Good	1.00	96	88	80	64	48	40	32
Regular	0.90	86	79	72	58	43	36	29
Poor	0.70	67	62	56	45	34	28	22
Very poor	0.40	38	35	32	26	19	16	13

Table A.5: *Fifteen Largest Industries (by Number of Firms in Sample)*

CNAE Code	Locally traded industries	CNAE Code	Nationally traded industries
1812	Apparel manufacturing, except underwear, sweaters, shirts	1931	Leather shoes
1581	Bakery products, confectionery and pastry manufacturing	2010	Wood splitting (logs)
2812	Manufacture of metal frames	2892	Artifacts manufacturing drawn
2222	Printing of educational material, and material for industrial or commercial uses	1779	Other textile products (knitting)
2842	Manufacturing locksmith articles, except frames	1939	Manufacture of footwear of other materials
2522	Manufacture of plastic packaging	2519	Manufacture of other rubber products
1589	Other food manufacturing	2833	Manufacturing metal stamped artifacts
2029	Manufacturing various wooden goods, straw, cork and braided materials, except furniture	2931	Manufacture of machinery and equipment for agriculture, poultry and obtaining animal products
1542	Dairy product manufacturing	2021	Plywood, laminated, pressed or sintered wood manufacturing
1813	Manufacture of professional apparel	2473	Manufacture of perfumes and cosmetics
2893	Manufacturing body shop items and metal articles for household use and personal	2940	Machine tool manufacturing
1749	Manufacture of other textile articles, including weaving	3210	Basic electronic material manufacturing
1543	Ice cream manufacturing	2481	Manufacture of paints, varnishes, enamels and lacquers
1584	Pasta manufacturing	2749	Other non-ferrous metals and alloy metallurgy
1921	Manufacture of bags, handbags, suitcases and other travel goods of any material	2221	Printing of newspapers, magazines and books

Source: Author's calculations based on RAIS data, year 2007. Definition of locally traded industries uses the Ellison-Glaeser index, with a value of 0.2 or less.

Nationally traded industries based on the Ellison-Glaeser index, with a value of 0.5 or more.

Table A.6: *Correlations Between the Key Variables*

	Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	Survive	1.00											
2	Relocate	-0.14	1.00										
3	Switch product	-0.28	0.06	1.00									
4	Change in co-location	-0.01	-0.01	0.01	1.00								
5	Number of workers (log)	0.14	0.07	0.09	-0.02	1.00							
6	Average annual worker wage (log)	0.09	0.04	0.05	-0.05	0.25	1.00						
7	Firm age in years (log)	0.11	-0.03	0.03	0.01	0.15	0.06	1.00					
8	Exporter	0.05	0.02	0.03	-0.03	0.30	0.15	0.11	1.00				
9	Importer	0.06	0.03	0.03	-0.03	0.32	0.18	0.10	0.48	1.00			
10	Baseline co-location (log)	0.01	-0.03	-0.11	-0.04	-0.01	-0.01	0.01	-0.03	-0.05	1.00		
11	Urbanization economies (log)	0.05	0.03	0.00	-0.03	0.08	0.22	0.12	0.09	0.13	0.51	1.00	
12	Strength of competition (log)	0.01	-0.01	-0.08	0.08	-0.23	-0.07	-0.01	-0.09	-0.11	0.30	0.22	1.00

Note: Own-calculations based on data from Relação Anual de Informações Sociais (RAIS), SECEX, and Ministry of Transport of Brazil.

A.2 Supplementary Figures for Chapter 1

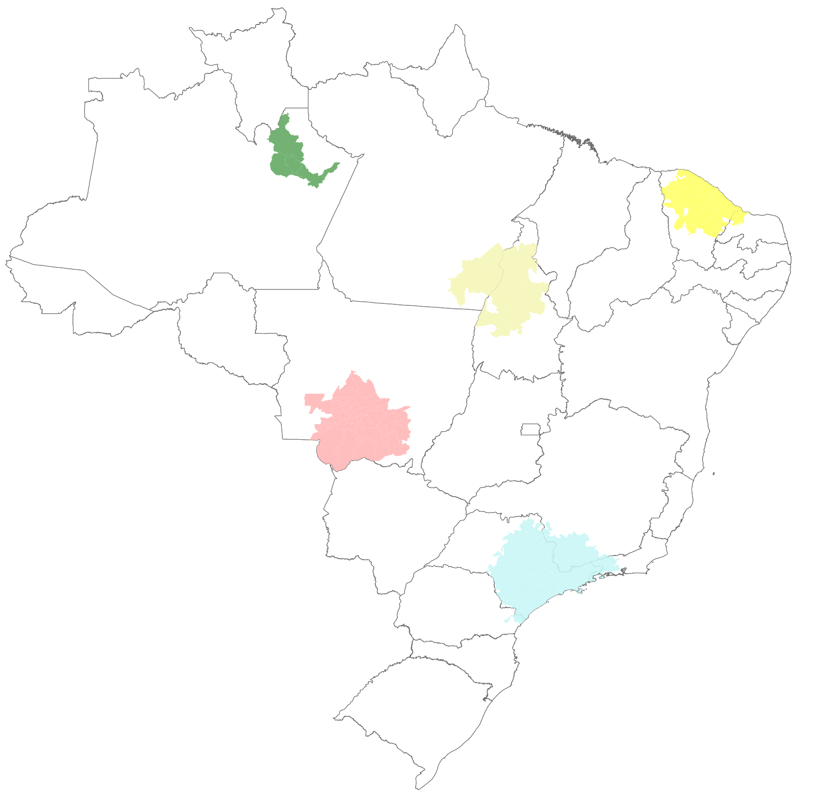

Figure A.1: *Examples of Markets as Defined by Four Hours of Travel Time*

Note: Author's calculations based on data from the Ministry of Transport of Brazil.

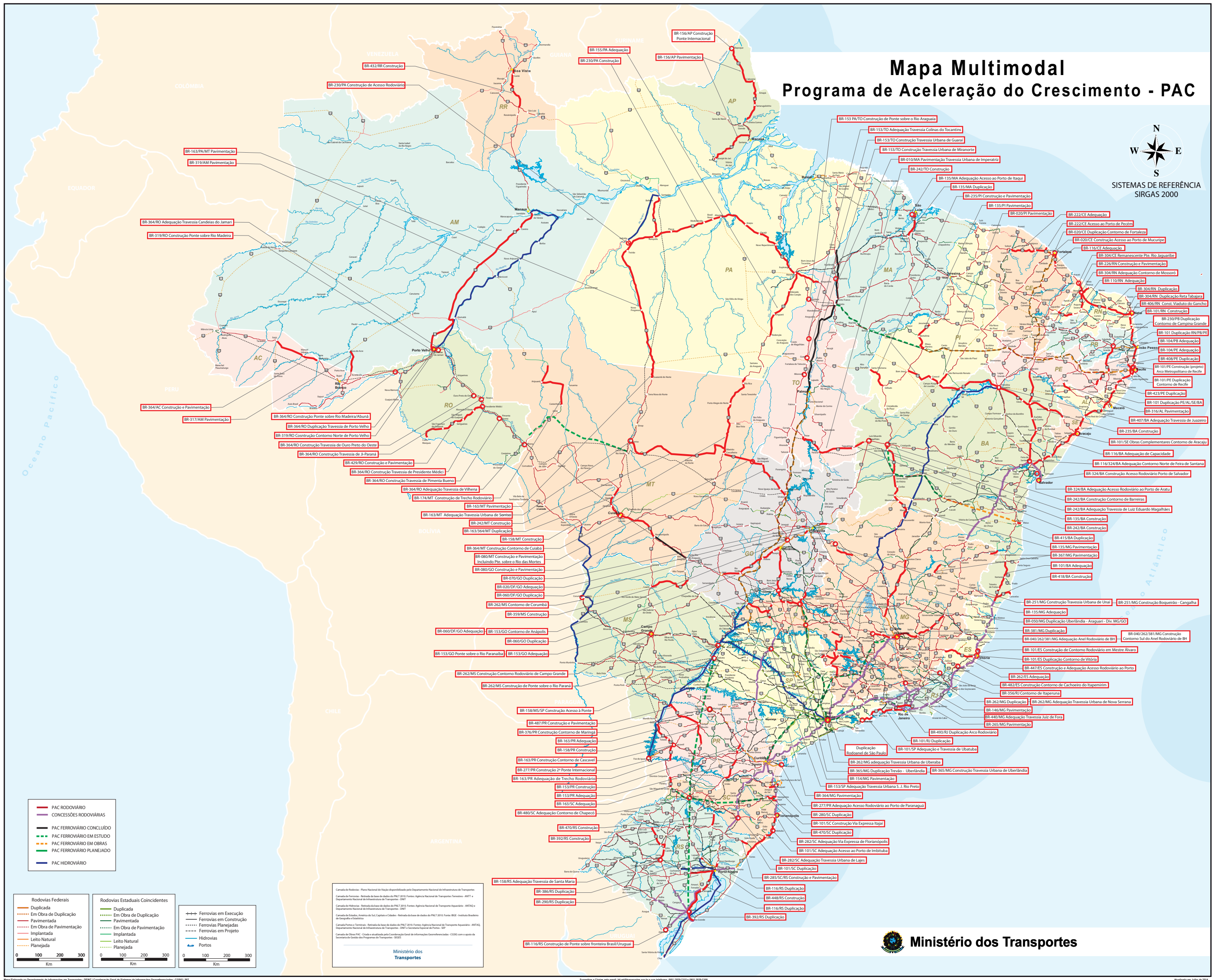

Figure A.2: *Programa de Aceleração do Crescimento — Programmed Investments*

Source: Ministerio dos Transportes de Brasil.

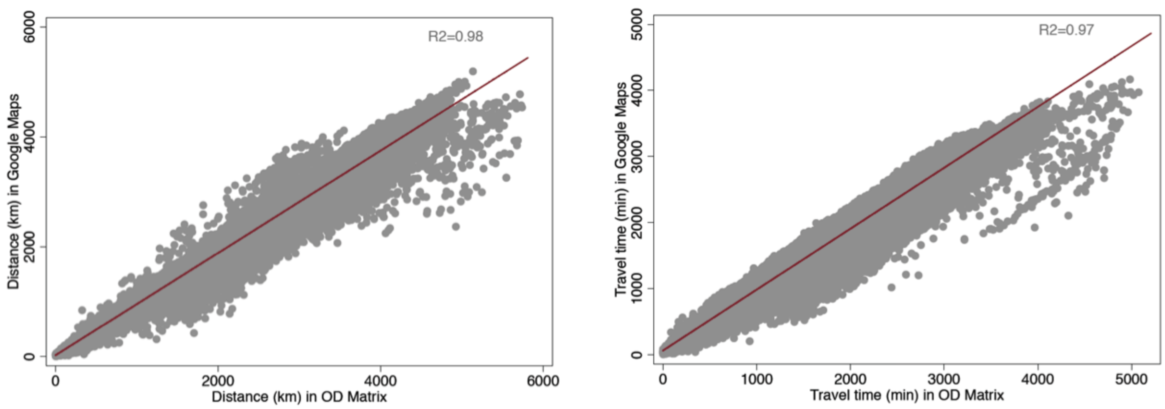

Figure A.3: *Distances (left) and Travel Times (right) in Google Maps and O-D matrix*

Note: Author's calculations based on data from the Ministry of Transport of Brazil and Google Maps.

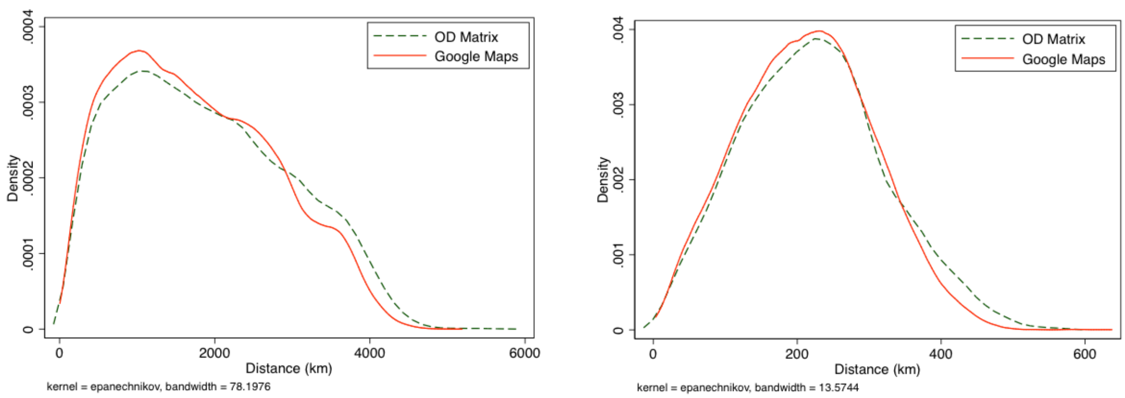

Figure A.4: *Distribution of Distances Globally (left) and Locally (right)*

Note: Author's calculations based on data from the Ministry of Transport of Brazil and Google Maps.

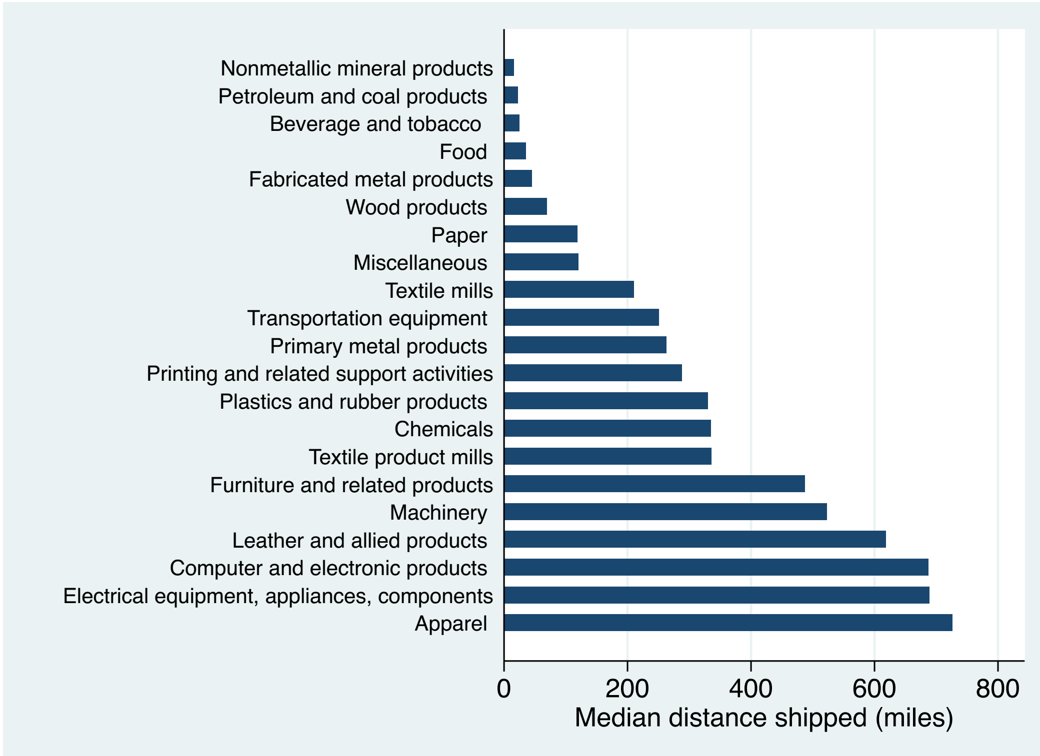

Figure A.5: *Median Distance Manufacturing Goods Shipped by Truck, by NAICS sector*

Note: Author's calculations using U.S. 2012 Commodity Flow Survey (CFS) data.

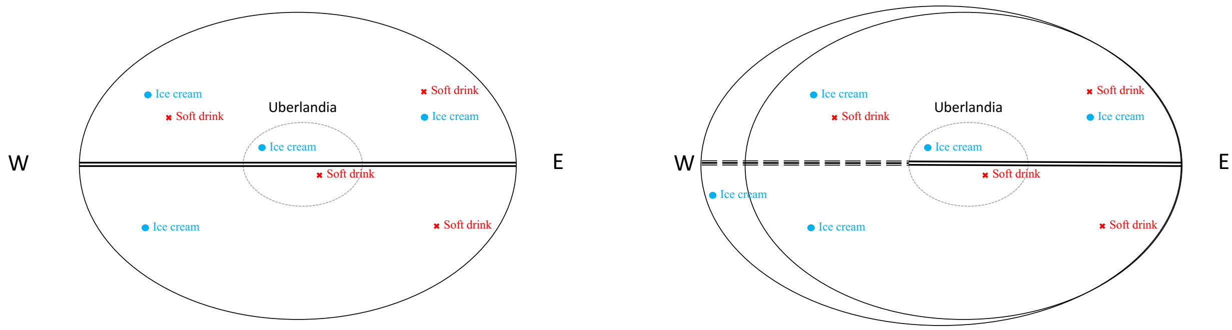

Figure A.6: *Example of Two Industries in Uberlandia Before (left) and After (right) Road Investment*

Note: The figure illustrates a hypothetical example of two firms located in the municipality of Uberlandia, one producing ice cream and the other soft drinks. The figure illustrates how the direction of a road investment i) increases the overall market size that lies within four hours of the focal firm and ii) changes the proximity of firms to their existing competitors. The main insight embedded in the figure is that the *pre-existing* locations of the firms in each industry combined with the direction of the road investment together determine the size of the change in co-location resulting from the road.

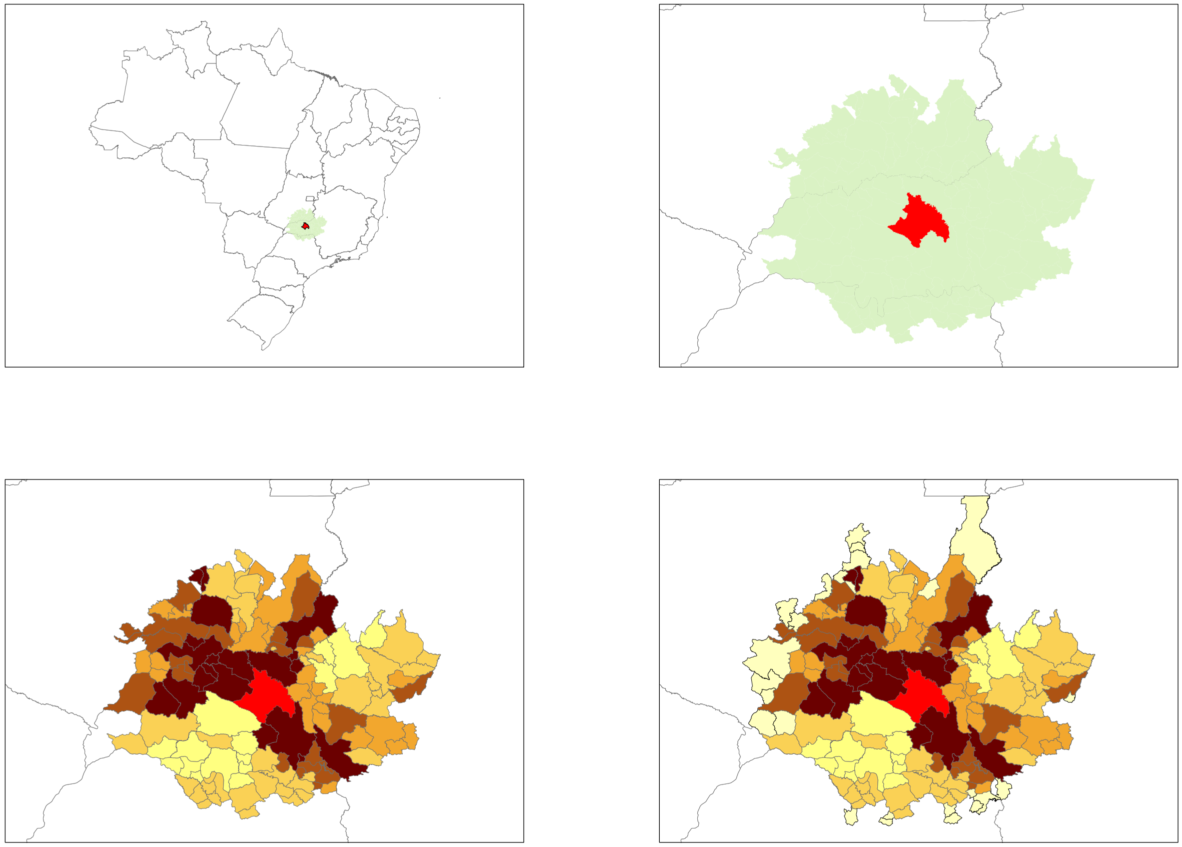

Figure A.7: *Variation in Local Travel Times in Uberlandia, 2007–2014*

Note: The Figure illustrates the location of Uberlandia in Brazil (top, left) and the extent of its four-hour market (top, right). Darker-shaded regions represent greater changes in travel time between Uberlandia and the municipality (bottom, left). Changes in travel times can also cause an expansion of the four-hour market area (bottom, right).

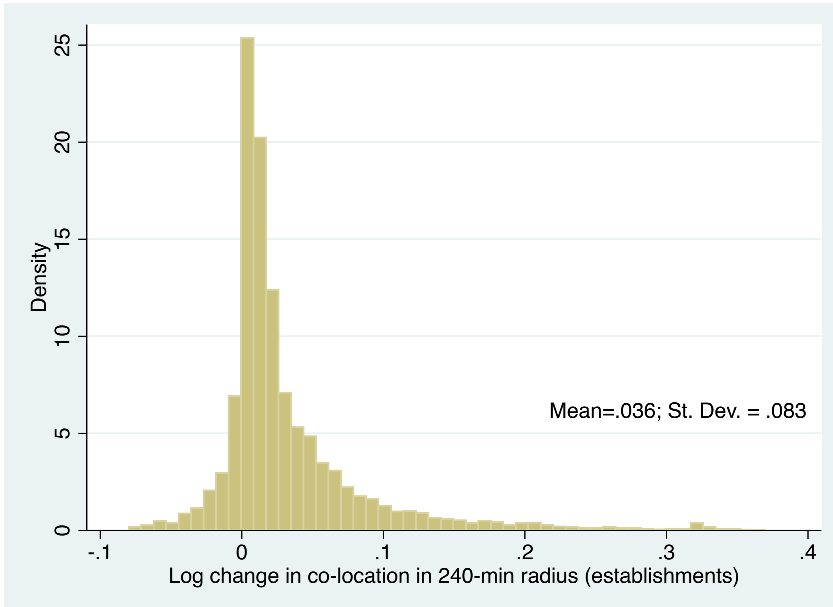

Figure A.8: *Distribution of Change in Co-Location*

Note: Author's calculations based on data from the Ministry of Transport of Brazil and RAIS.

A.3 Supplementary Tables for Chapter 2

Table A.7: *Most Agglomerated Industry Pairs, Multibusiness (top) and Stand-alone (bottom)*

Rank	Industry 1	SIC3	Indistry2	SIC3	Agg. at 250 km.
1	Broadwoven fabric mills, cotton	221	Household furniture	251	0.872
2	Pulp mills	261	Newspapers	271	0.753
3	Broadwoven fabric mills, cotton	221	Paperboard mills	263	0.680
4	Broadwoven fabric mills, wool	223	Textile finishing, except wool	226	0.679
5	Broadwoven fabric mills, cotton	221	Narrow fabric mills	224	0.657
6	Broadwoven fabric mills, cotton	221	Paperboard containers and boxes	265	0.640
7	Broadwoven fabric mills, cotton	221	Textile finishing, except wool	226	0.626
8	Broadwoven fabric mills, synthetics	222	Broadwoven fabric mills, wool	223	0.559
9	Broadwoven fabric mills, wool	223	Yarn and thread mills	228	0.559
10	Broadwoven fabric mills, cotton	221	Broadwoven fabric mills, wool	223	0.543
11	Broadwoven fabric mills, cotton	221	Industrial inorganic chemicals	281	0.541
12	Misc. wood products	249	Household furniture	251	0.532
13	Narrow fabric mills	224	Yarn and thread mills	228	0.525
14	Textile finishing, except wool	226	Yarn and thread mills	228	0.521
15	Broadwoven fabric mills, synthetics	222	Textile finishing, except wool	226	0.519
1	Women's and misses' outerwear	233	Girls' and children's outerwear	236	0.147
2	Women's and misses' outerwear	233	Women's and children's undergarments	234	0.140
3	Knitting mills	225	Women's and misses' outerwear	233	0.140
4	Men's and boys' furnishings	232	Women's and misses' outerwear	233	0.138
5	Broadwoven fabric mills, wool	223	Women's and misses' outerwear	233	0.137
6	Broadwoven fabric mills, wool	223	Knitting mills	225	0.135
7	Women's and misses' outerwear	233	Handbags and personal leather goods	317	0.135
8	Women's and misses' outerwear	233	Jewelry, silverware, and plated ware	391	0.131
9	Knitting mills	225	Yarn and thread mills	228	0.129
10	Women's and misses' outerwear	233	Fur goods and misc. apparel and accessories	238	0.126
11	Men's and boys' suits and coats	231	Women's and misses' outerwear	233	0.125
12	Knitting mills	225	Girls' and children's outerwear	236	0.124
13	Knitting mills	225	Men's and boys' suits and coats	231	0.123
14	Knitting mills	225	Women's and children's undergarments	234	0.122
15	Women's and misses' outerwear	233	Costumes & Miscellaneous	396	0.122

A.4 Supplementary Figures for Chapter 2

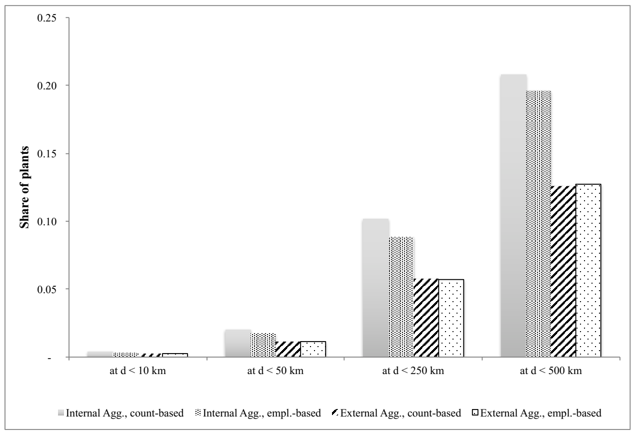

Figure A.9: *Average Internal and External Agglomeration Across Industry Pairs. Count- and Employment-Based Measures*

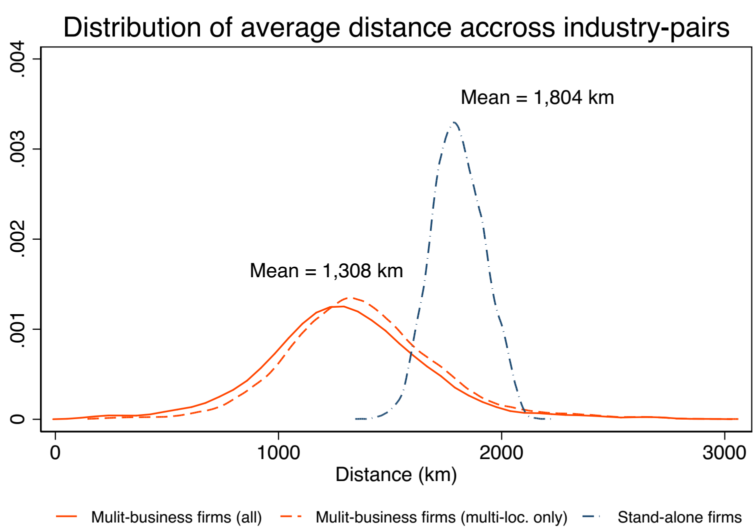

Figure A.10: *Average Distance Across Industry Pairs.*

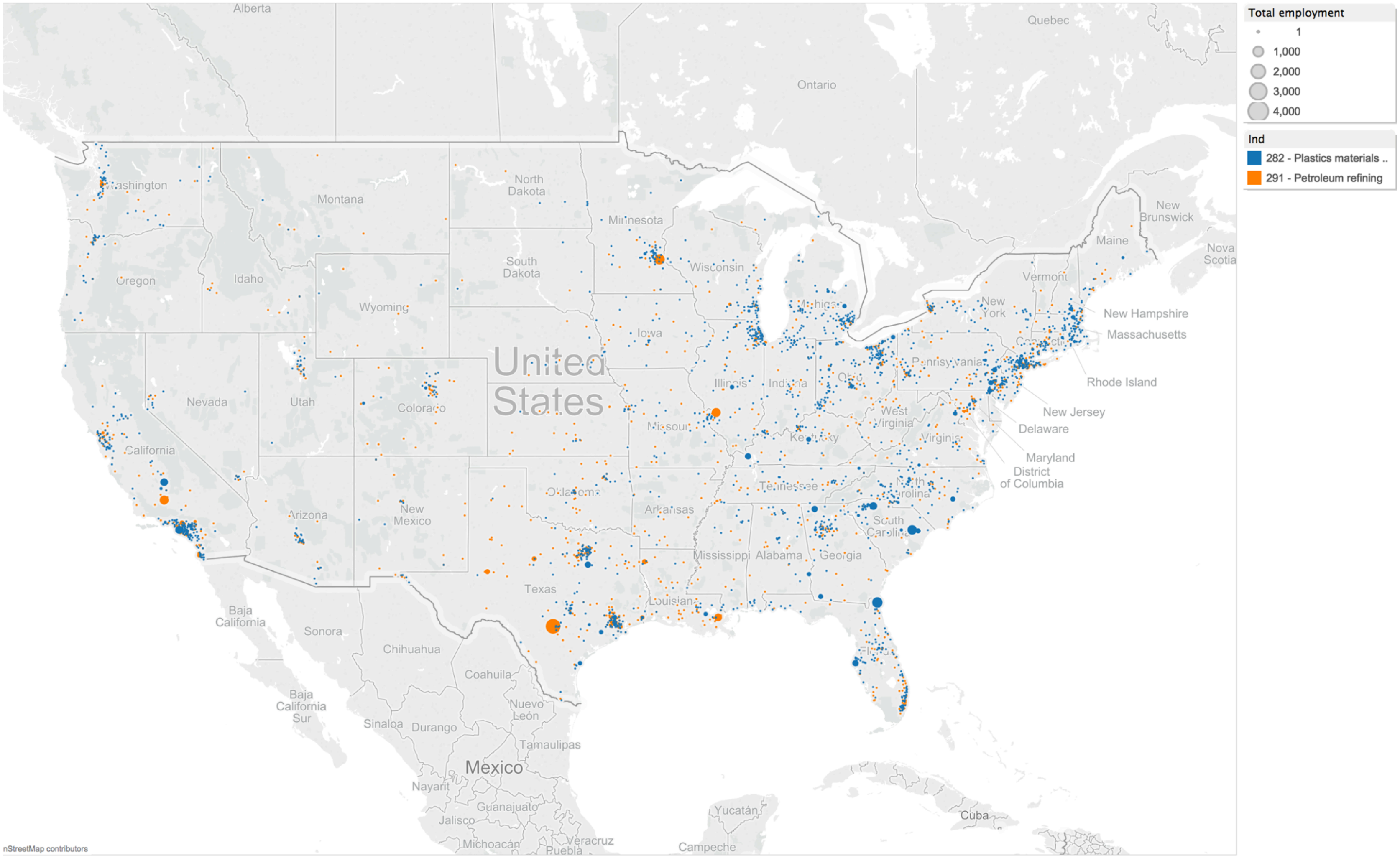

Figure A.11: *Location of All Stand-Alone Firms in Plastics Materials (SIC 282) and Petroleum Refining (SIC 291).*

Note: Authors' illustration using data from Dun & Bradstreet, 2012.

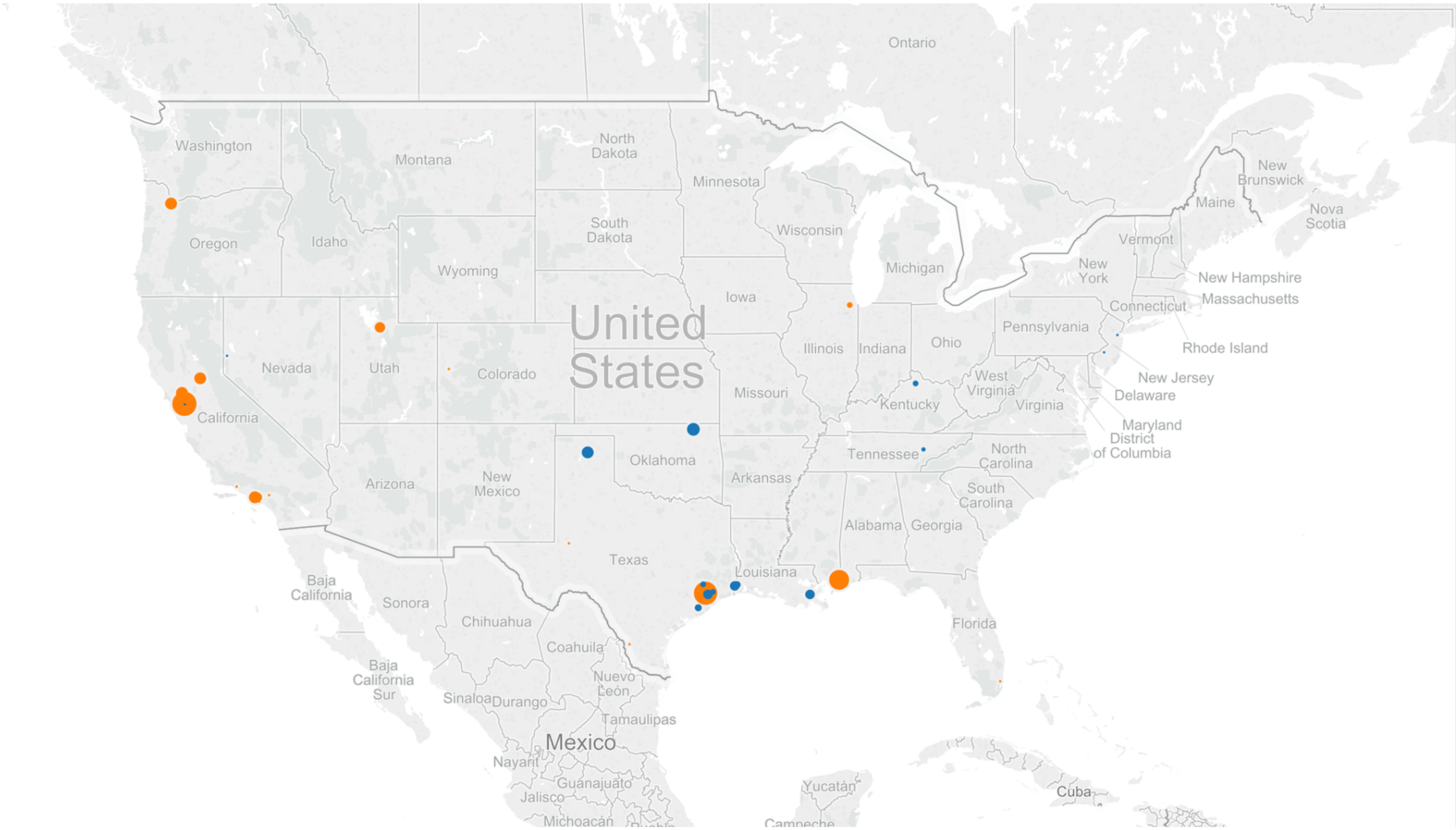

Figure A.12: *Location of Plants in Plastics Materials (SIC 282) and Petroleum Refining (SIC 291) of one Multi-Business Firm.*

Note: Authors' illustration using data from Dun & Bradstreet, 2012.

A.5 Supplementary Figures for Chapter 3

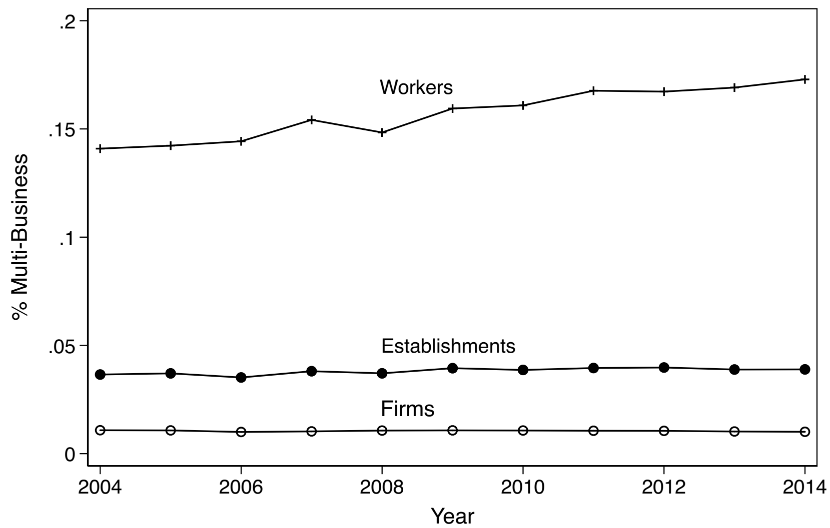

Figure A.13: *Share of Multibusiness Firms in all Firms, Establishments, and Employment, Brazil 2004–2014.*

Note: Authors' calculations using data from RAIS.