Title
Essays in international trade and entrepreneurship

Author
Hirakawa, Oana

Publication Date
2011

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays in International Trade and Entrepreneurship

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

in

Economics

by

Oana Hirakawa

Committee in charge:

Professor Gordon Hanson, Chair
Professor Thomas Baranga
Professor Eli Berman
Professor Lawrence Broz
Professor Roger Gordon

2011

The dissertation of Oana Hirakawa is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2011

DEDICATION

To wonderful Nina.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Gordon Hanson for his support as the chair of my committee. His insightful feedback, patient guidance, and constant encouragement have proven invaluable to completing my thesis.

I also thank the members of my committee Thomas Baranga, Eli Berman, Lawrence Broz, and Roger Gordon for their helpful feedback. Special thanks go to my husband Brad Hirakawa for stubbornly reminding me that there's more to life than work and school, my parents Felicia and Ioan Tocoian for always supporting my choices, and my classmates and friends at UCSD, who made even the more challenging times in graduate school seem fun.

Chapter 1 has benefitted from comments from Gordon Hanson, Sarada, Thomas Baranga, Eli Berman, Lawrence Broz, Tai Ming Cheung, Tiffany Chou, Daniel Egel, Roger Gordon, Youjin Hahn, Craig McIntosh, Marc Muendler, Karthik Muralidharan, Philip Neary, Paul Niehaus, James Rauch, Hee Seung Yang, Li Zhou, and seminar participants at the UCSD Economics Department and IR/PS.

I thank Gordon Hanson, Thomas Baranga, Eli Berman, Lawrence Broz, Roger Gordon, and Chad Krause for comments and suggestions regarding chapter 2.

Chapter 3 is currently being prepared for submission for publication of the material. Sarada; Hirakawa, Oana. We thank Roger Gordon, Gordon Hanson, Sam Bazzi, Silke Forbes, Craig McIntosh, Philip Neary and Chris Woodruff for many helpful comments. We also thank Marc Muendler for providing us with access to the RAIS data.

Chapter 4, in full, has been submitted for publication of the material as it may appear in the International Journal of Industrial Organization. Hirakawa, Oana; Muendler, Marc-Andreas; Rauch, James E. We have benefited from comments by Victor Gomes, John Haltiwanger, Daniel Lederman, the editor and referees, and seminar participants at Georgia Tech, New York University, Universidade de Brasília, UC Davis, and the World Bank. We acknowledge support from the International Growth Centre (RA-2010-03-002) and the World Bank with gratitude.

VITA

2003	B. S. in Engineering and Applied Science, California Institute of Technology
2007	M. S. in Economics, University of California, San Diego
2011	Ph. D. in Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Essays in International Trade and Entrepreneurship

by

Oana Hirakawa

Doctor of Philosophy in Economics

University of California, San Diego, 2011

Professor Gordon Hanson, Chair

This dissertation contains four chapters of my research.

In chapter 1, I show that higher military spending leads to higher exports of weapons. This is a manifestation of the home market effect - the prediction that countries with higher demand for a differentiated good will be net exporters of that good. The home market effect is specific to monopolistic competition models of international trade, and serves to distinguish empirically between these and comparative advantage-based models. I construct a monopolistic competition model with a military sector and a continuum of civilian industries, and derive empirical implications of the home market effect for arms and ammunition. I then use military expenditure as a measure of demand for military weapons to show that, indeed, countries with higher military spending as a

share of GDP export more arms and ammunition relative to homogeneous, cheap-to-ship civilian goods. In my setup, military spending serves to introduce variation in demand across countries. This is an innovation over the typical approach in the literature, whereby consumers in different countries are assumed to have identical preferences, and home market effects stem from differences in goods' characteristics and country size alone.

Chapter 2 tests the existence of the home market effect for construction materials, by using public infrastructure spending as a measure of demand. I construct a theoretical model that suggests goods with high transport costs and high differentiation are most likely to display home market effects. I test this prediction empirically for a handful of construction materials that meet the necessary criteria. As expected, I find that the home market effect holds for alloy steel and construction machinery. However, cement and glass display the opposite trade pattern, whereby increased domestic demand leads to reduced exports. I discuss potential explanations for this result.

In the next two chapters, I examine the determinants of firm success. In chapter 3 (co-authored with Sarada) we find that new firms with higher network concentration, i.e. wherein initial employees have worked together previously, are on average larger, have higher wages and survive longer. This association increases with the initial size of the newly founded firm. However, we find a negative relationship between network concentration and initial firm growth. Finally, chapter 4 (co-authored with Marc Muendler and James Rauch) gauges the prevalence and performance of firms founded as employee spinoffs relative to other entrants. We find that size at entry is larger for employee spinoffs than for new firms without parents but smaller than for diversification ventures of existing firms. Similarly, survival rates for employee spinoffs are higher than for new firms without parents and comparable to those for diversification ventures of existing firms.

Chapter 1

The Home Market Effect and the International Arms Trade

Abstract

In a novel mechanism of international arms proliferation, I show that higher military spending leads to higher exports of weapons. This is a manifestation of the home market effect - the prediction that countries with higher demand for a differentiated good will be net exporters of that good. The home market effect is specific to monopolistic competition models of international trade, and serves to distinguish empirically between these and comparative advantage-based models. I construct a monopolistic competition model with a military sector and a continuum of civilian industries, and derive empirical implications of the home market effect for arms and ammunition. I then use military expenditure as a measure of demand for military weapons to show that, indeed, countries with higher military spending as a share of GDP export more arms and ammunition relative to homogeneous, cheap-to-ship civilian goods. In my setup, military spending serves to introduce variation in demand across countries. This is an innovation over the typical approach in the literature, whereby consumers in different countries are assumed to have identical preferences, and home market effects stem from differences in goods' characteristics and country size alone. An instrumentation strategy illustrates that the home market effect in the defense industry is sufficiently strong that countries suffering from conflict, and thereby expanding military expenditures, see their arms exports increase more than proportionately.

1.1 Introduction

Traditional models of international trade predict that trade is driven by differences in factor endowments or production technologies, with each country exporting goods in which it has a comparative advantage. If we allow for variation in demand in this setting, we find that countries with higher demand for a good are net importers of that good.[1] In the monopolistic competition model, Paul Krugman introduces scale economies and draws predictions about the direction of trade from differences in demand: differentiated increasing returns to scale goods will be produced in, and exported from, the market that has higher demand for them.[2] This is the home market effect, and it can be used to distinguish empirically between the two classes of models discussed: comparative advantage vs. monopolistic competition models.

In practice, most goods do not have a reliable measure of demand - one that is set independently of prices. Therefore the typical approach to the home market effect has been to assume that individual consumers have identical preferences, and simply use gross domestic product to measure demand: in a pair of countries, the larger one is said to have higher demand for all goods; the home market effect is then the prediction that the large country exports more differentiated goods, while the small country exports more constant returns to scale, homogeneous goods. Among the few exceptions, Davis and Weinstein (1999) look at how Japanese regional variation in demand influences production of goods within industries, and find that home market effects matter in several manufacturing industries. Davis and Weinstein (2003) run a similar analysis for OECD countries, and again find that a significant number of sectors display home market effects. However, their absorption measure offers unbiased estimates only under the assumption that industry demand shocks are uncorrelated with industry supply shocks.

In this paper, I use government military spending to measure the demand for arms and ammunition, which allows me to introduce variation in the patterns of demand across countries. Results are no longer dependent on the absolute size of nations - I use relative demand for military vs. civilian goods, and test that countries with higher military spending as a share of GDP export more arms relative to homogeneous civilian goods. This represents a closer interpretation of the original home market effect

[1] This is true as long as comparative advantage patterns are not positively correlated with demand.

[2] For this result to obtain, an additional condition is needed: that these goods have non-negligible transport costs, since otherwise production could concentrate in any country, even one with zero consumption of the good in question.

formulation, and a more direct empirical verification of it.

Within the arms industry, my analysis leads to a surprising result: the more a country spends on its military, the more we might expect it to reduce exports of weapons due to increased consumption. Instead, I find that higher military spending is associated with higher arms exports. This implies a new mechanism for proliferation, through investments in domestic arsenals. Why do countries spend on weapons procurement? They may fear imminent wars and want to be able to defend themselves. They may want to amass an impressive arsenal in order to deter attacks. They may purposefully want to strengthen the domestic arms industry in order to have self-sufficiency, should trade lines be cut by conflict. But do countries expect that a surge in their military expenditure will lead to increased arms exports? Is military spending recognized as a source of outgoing arms proliferation? From a political science and international relations point of view, this is an unexpected and interesting result.

Going back to the empirical trade literature, standard treatments of the home-market effect (as in Helpman and Krugman, 1985; Feenstra, 2004) are based on a monopolistic-competition model of trade that has one homogeneous, constant returns-to-scale industry with zero transport costs, and one differentiated increasing-returns-to-scale industry with positive transport costs. Hanson and Xiang (2004) extend this model to allow for a continuum of differentiated-product industries.[3] They show that industries with high transport costs and low substitution elasticities (i.e., more product differentiation) tend to concentrate in the larger country, while industries with low transport costs and high substitution elasticities (i.e., less product differentiation) tend to concentrate in the smaller country. As in the standard treatment, they exploit differences in goods' characteristics and GDP to demonstrate the home market effect.

However, this approach has its drawbacks. Gross domestic product is only a

[3]Holmes and Stevens (2005) also consider a continuum of industries, allowing for variation in returns to scale, in a model that departs from the monopolistic competition framework. They predict that goods with very strong economies of scale will be produced in the large country, and that goods with weak economies of scale are non-traded.

loose approximation of demand for most goods, since preferences are of course not identical across countries.

I add a differentiated military sector in addition to the continuum of civilian industries from Hanson and Xiang (2004). This allows me to introduce variation in the cross-country patterns of demand by employing military expenditure as a measure of demand for arms and ammunition. I can then show the home market effect stems from both differences in goods' characteristics and differences in demand: I compare exports of arms and ammunition with exports of *control* goods (undifferentiated, cheap to transport), and with exports of *similar* goods (of substitution elasticity and freight rate close to those of arms). In both cases, I find that countries with higher military spending relative to GDP export significantly more arms. The effect is stronger when comparing to *control* rather than *similar* goods, as predicted by theory.

To further investigate the role of differentiation for the home market effect, I consider two sub-samples of the arms and ammunition product group: one I expect to be more differentiated (armored vehicles, rifles, machine guns, bombs and torpedoes, etc.) and one with more homogeneous goods (cartridges). I find that, for two out of the three samples of exporters considered, the home market effect is stronger for the first product subsample (results are identical for the other sample of exporters).

There are a number of possible concerns about drawing inference from the military arms sector, the most notable of which is that economic principles may take a back seat to military strategy in the arms trade. How much does military strategy dictate the arms trade? The political science literature suggests that post-Cold War, the international trade in arms has been driven increasingly by economic considerations, to the detriment of military strategy considerations (see for example Lumpe, 1999). This means the arms sector behaves similarly to civilian industries, and we are justified in employing a conventional international trade model to describe the trade in arms. However, concerns about the importance of military interests may linger, therefore I control for NATO membership in the empirical analysis: I only compare pairs of countries that have the same membership status.

My identification strategy relies on the assumption that military expenditure is independent of economic factors which might influence arms exports. I argue that military spending is driven primarily by idiosyncratic preferences and perceived threats, and that, due to institutional constraints, it is very persistent over time. This limited responsiveness makes it exogenous and therefore a suitable demand measure. Nonetheless, I also run instrumental variables regressions with two different types of instruments, in order to address remaining concerns.

Several authors have previously examined the determinants of military spending: Hewitt (1991) seeks to explain cross-country variation in military expenditure in a panel dataset of 125 countries during 1972-88. He finds that the ratio of military expenditure and GDP increases moderately with GDP, but only at low levels of income. He also finds that involvement in a conflict, form of government and geographical variables matter. Sandler and Hartley (1995) model the demand for military expenditure empirically as influenced by GDP, allies' and enemies' expenditures, and taking part in a conflict, but estimate the model on a country-by-country basis, thus acknowledging that country-specific preferences explain much of the variation. Chowdhury (1991) uses Granger tests, but finds no general causal relationship between GNP and military spending as a share of GNP across countries. Several other studies also fail to find a clear pattern of Granger causality between economic variables and military spending, as noted by Smith (1995), who also demonstrates the persistent nature of military spending in country-by-country regressions.

I find that military spending is strongly persistent, with nearly 99 percent of variation in a 20-year panel of high income OECD countries explained by the previous year's value. It is also highly country-specific: 95 percent of variation is explained by country and year fixed effects alone. Finally, in order to quell remaining concerns about potential endogeneity of military expenditure, I estimate the model using an instrumental variable approach: I use five and ten-year lags of military spending to instrument for current procurement, thus eliminating potential contemporaneous reverse causality or omitted variable bias. I find results are the same as when using ordinary least squares estimation, which justifies the assumption that military expenditure is an exogenous demand measure in this context. Nonetheless, this is not a purely cross-section analysis: the data do have time series variation as well: I show that even the positive variation in military expenditure due to conflicts leads to higher arms exports. This goes against typical expectations and showcases the strength of the home market effect.

I show that arms production expands more than proportionately with military expenditure, and claim this is due to economies of scale. Could the result be driven instead by another mechanism? Suppose large military powers intentionally over-produce and export arms during peacetime, in order to maintain excess capacity they can appeal to during conflicts - this would lead to the pattern observed even in the absence of scale economies. Gold (1999) weighs the two possible mechanisms behind expanding military production. First, he lists economic motivations: military production is characterized by both economies of scale due to high R&D and capital investment costs, and by learning economies, both of which justify exports as a way to lower unit costs. He then mentions the excess capacity motivation: strong exports means production lines don't have to be closed, so they are available immediately in case of sudden need. In addition, a surge in production can be setup fairly easily. However, he immediately questions whether maintaining production lines is cheaper than restarting, and mentions that during the Gulf War there was no need for excess capacity. By incorporating data on international conflicts, I can directly test and reject the latter mechanism: I find that not only do arms exports not drop during conflicts, but conflicts actually lead to higher arms exports through increased military spending. For some samples of exporters and comparison goods, arms exports increase during conflicts even independently of rising military spending.

A similar alternate mechanism suggests that countries may make large long-term investments in the military sector in order to export arms to allies. However, this is an unlikely motivation, since in practice investments in arms production are justified either by the need to be self-sufficient, or by expected economic gains. Clearly, the economic gains incentive is consistent with my story. The self-sufficiency motivation is also compatible, since it is through economies of scale that production which was initially set up to meet domestic demand expands into exports.

Yet another potentially confounding issue is military aid: military expenditure data includes aid, most of which is given in the form of credits to purchase arms from the donor country. This can lead to a mechanical link between high military spending and high arms exports. However, even for big donor countries, military aid represents a small minority of military spending.[4] In addition, results are robust when I exclude top military aid donors.

The rest of the paper is organized as follows: section 1.2 presents a theoretical model that incorporates a civilian and a military sector, and derives empirical implications of the home market effect in this setting; section 1.3 discusses data considerations, section 1.4 performs empirical tests of the home market effect, and section 1.5 concludes.

1.2 Theoretical model setup

I model two types of goods: a continuum of differentiated civilian industries, whose products are demanded by consumers, and a differentiated military sector, whose goods are demanded by the government exclusively.[5]

There is a large country and a small country. Each has one factor of production: labor. The large country has a mass $L > 1$ of workers, each earning wage w. The small country's labor endowment and wage are normalized to 1 (so $w^*L^* = 1$). Each country's military budget ME (ME^*) is extracted from workers' income by lump-sum taxation,

[4]As of 2006-2008, the United States disbursed approximately $4.5 billion per year through its Foreign Military Financing program, making it by far the largest military aid donor. This amount represents less than one percent of American military expenditure.

[5]In reality, of course, national defense enters the utility function of individual consumers, but I assume that once government has been elected, it is an independent player whose sole and exclusive domain is to acquire goods necessary for defense. Military armament is then treated as a private rather than public good, simplifying the model.

so that workers will have after-tax income $Y = wL - ME$ ($Y^* = 1 - ME^*$) to spend on civilian goods, while governments spend ME (ME^*) on military goods.

1.2.1 Civilian goods industries

Civilian goods are modeled on a continuum, in order to allow for variation in differentiation and transport costs - the two dimensions that will determine which industries display home market effects. In particular, I consider a continuum of monopolistically competitive industries (as introduced by Dixit and Stiglitz, 1977) indexed by $z \in [0,1]$. Consumers derive utility from purchasing many different varieties of a given product.[6] Each variety is characterized by increasing returns to scale, so in equilibrium it will be produced by a single firm. Firms continue to enter until the last firm just breaks even. Since cost structures are identical across firms, in equilibrium all firms have zero profits.

First, I outline the consumers' problem: individuals have Cobb-Douglas preferences over industries, and constant elasticity of substitution (CES) demand over varieties within an industry:

$$U_{\text{consumer}} = \prod_{z \in [0,1]} \left[\left(\sum_{i=1}^{n(z)} q_{zi}^{\frac{\sigma(z)-1}{\sigma(z)}} \right)^{\frac{\sigma(z)}{\sigma(z)-1}} \right]^{\alpha(z)}$$

In the equation above, $\alpha(z)$ is the consumption share of industry z products and $\int_0^1 \alpha(z)dz = 1$; $n(z)$ is the number of product varieties in industry z, $\sigma(z)$ is the elasticity of substitution between varieties (restricted to be larger than one), and q_{zi} is the quantity of variety i in industry z.

Let $\tau(z) > 1$ be the iceberg transport cost incurred in shipping one unit of output

[6]This may make sense at the individual level if the product is food or shoes, but do people really purchase a little bit of every type of car? No, but in this case aggregation saves the argument: each consumer may only purchase one car, but their friends and neighbors will want to differentiate themselves by buying a different brand or model, so once we've aggregated up to the region or country level, consumption patterns are consistent with the love-of-variety approach.

from one country to the other, and $x(z) = \tau(z)^{\sigma(z)-1}$ the effective trade cost[7] for industry z.

I will assume there is no international specialization at the industry level, meaning each country produces some goods in each industry. The varieties of industry z are symmetric: let $c(z)$ be the fixed labor requirement, and I normalize the variable labor requirement for each variety to one. Then output and price are the same for all varieties: $q_{zi} = q(z)$, $p_{zi} = p(z)$. As a result of the CES demand specification, the price is a constant markup over marginal cost (in this case, wage w):

$$p(z) = \frac{\sigma(z)}{\sigma(z)-1}w \tag{1.1}$$

Since free entry drives profits to zero, output is fixed and revenues are proportional to fixed costs: $\Pi(z) = p(z)q(z) - [c(z)w + qw] = 0$, and we replace the expression for $p(z)$ from equation (2.1) to find:

$$\begin{aligned} q(z) &= c(z)[\sigma(z)-1] \\ p(z)q(z) &= wc(z)\sigma(z) \end{aligned}$$

1.2.2 Military goods industry

In deciding how to model demand for military goods, I considered the fact that modern war is multifaceted, and a nation that wishes to defend itself against unknown future threats has to be ready to operate in a variety of battle theaters, using a synergy of weapons. For example, the United States Armed Forces are composed of five separate service branches: Army, Navy, Marine Corps, Air Force, and Coast Guard, each with its own designated area of operations and specialized arsenal. And while there are some common staples, like the M16 rifle, there is also remarkable diversity in the range of weapons employed within and across branches, from submachine guns, to light and

[7]As in all monopolistic competition models, transport costs matter more for industries with high elasticity of substitution. The exact specification of x will become obvious shortly in the model derivation.

heavy machine guns, grenades, rockets, missiles and their launching systems, unmanned vehicles, armored trucks, tanks, helicopters, fighter jets, etc.

I therefore consider the love-of-variety approach to be suited for the arms sector as well, and I use the monopolistic competition model with CES aggregator to represent in reduced form the government's decision over arms purchases. Mathematically, the military goods industry will be characterized by the same variables as any individual civilian industry z. To distinguish military goods, I mark their variables by subscript m.

$$\begin{aligned} U_{\text{government}} &= \left(\sum_{i=1}^{n_m} q_{mi}^{\frac{\sigma_m-1}{\sigma_m}}\right)^{\frac{\sigma_m}{\sigma_m-1}} \\ \Rightarrow p_m &= \frac{\sigma_m}{\sigma_m-1}w \\ q_{mi}=q_m &= c_m[\sigma_m-1] \\ p_m q_m &= w c_m \sigma_m \end{aligned}$$

1.2.3 Trade equilibrium

Let Γ be the share of after-taxes income spent by domestic consumers on domestic (civilian) goods, and Γ^* the share of income spent by foreign consumers on domestic goods.

Then the market for each civilian industry z product clears (z's left out for convenience):

$$npq = \alpha Y\Gamma + \alpha Y^*\Gamma^* \quad (1.2)$$

$$n^*p^*q = \alpha Y(1-\Gamma) + \alpha Y^*(1-\Gamma^*) \quad (1.3)$$

where
$$\Gamma = \frac{np^{1-\sigma}}{np^{1-\sigma}+n^*(\tau p^*)^{1-\sigma}} = \frac{np^{1-\sigma}}{np^{1-\sigma}+n^*(p^*)^{1-\sigma}x^{-1}}$$

$$\Gamma^* = \frac{n(\tau p)^{1-\sigma}}{n(\tau p)^{1-\sigma}+n^*(p^*)^{1-\sigma}} = \frac{np^{1-\sigma}}{np^{1-\sigma}+n^*(p^*)^{1-\sigma}x}$$

Military goods' market clears:

$$
\begin{aligned}
n_m p_m q_m &= ME\Gamma_m + ME^*\Gamma_m^* \\
n_m^* p_m^* q_m &= ME(1-\Gamma_m) + ME^*(1-\Gamma_m^*) \\
&\Gamma_m = \frac{n_m p_m^{1-\sigma_m}}{n_m p_m^{1-\sigma_m} + n_m^*(p_m^*)^{1-\sigma_m} x_m^{-1}} \\
&\Gamma_m^* = \frac{n_m p_m^{1-\sigma_m}}{n_m p_m^{1-\sigma_m} + n_m^*(p_m^*)^{1-\sigma_m} x_m}
\end{aligned}
$$

I arrive at the following equilibrium condition (see appendix 1.8.1 for derivation details):

$$
\begin{aligned}
0 &= \int_0^1 \alpha(z)g(z)dz + g_m && (1.4) \\
\text{where} \quad g(z) &= \left[\frac{Y}{x(z)w^{\sigma(z)}-1} - \frac{Y^*}{x(z)w^{-\sigma(z)}-1}\right] && (1.5) \\
g_m &= \left[\frac{ME}{x_m w^{\sigma_m}-1} - \frac{ME^*}{x_m w^{-\sigma_m}-1}\right] && (1.6)
\end{aligned}
$$

Both $g(z)$ and g_m are strictly decreasing in w, so equation (2.2) has a unique solution $w > 1$, as long as $\left[(Y-Y^*)\int_0^1 \frac{\alpha(z)dz}{x(z)-1} + (ME-ME^*)\frac{1}{x_m-1}\right] > 0$, a sufficient condition for which is that both the civilian and military sectors of the big country are larger than those of the small country. (see appendix 1.8.1 for the proof).

In the next section I will show that functions $g(z)$ and g_m code the trade-offs in the strategic decision over location faced by firms and, they are the key to whether a certain industry displays home market effects or not. Appendix section 1.8.1 contains comparative statics analysis.

1.2.4 Home market effect (HME)

A typical formulation for HME (as in Hanson and Xiang, 2004) is that industry z displays home market effects if the large country's share of varieties of z produced globally exceeds its share of world factor supplies; however, this is after assuming an identical demand structure across countries, which does not apply here.

Going back to the classic (Krugman, 1980) formulation, the home market effect arises from differences in demand, aside from (or instead of) country size. In the present paper, as in Krugman, preferences are not identical: demand *within* the civilian and military sectors follows the same pattern across countries, but governments idiosyncratically dictate how income is allocated *across* these two sectors. Therefore I develop a more general definition on the home market effect.

Definition - home market effect

Industry z is said to show home market effects if the country with higher demand for z produces a larger share of world z output than its share of world demand for z. In my 2-country world, that translates to:

a) For civilian industries indexed by z: $\frac{n(z)p(z)q(z)}{n^*(z)p^*(z)q(z)} = \frac{n(z)w}{n^*(z)} > \frac{\alpha(z)Y}{\alpha(z)Y^*} = \frac{Y}{Y^*}$.

Define $\tilde{n}(z) = n(z)w$. Then the condition is $\tilde{n}(z)/\tilde{n}^*(z) > \frac{Y}{Y^*}$ or $n(z)/n^*(z) > \frac{Y/w}{Y^*}$.

b) Under the assumption that the larger country (Home) also has higher military expenditure ($ME > ME^*$)[8], the military sector displays the home market effect if and only if $\tilde{n}_m/\tilde{n}_m^* > \frac{ME}{ME^*} \Leftrightarrow n_m/n_m^* > \frac{ME/w}{ME^*}$.

In terms of the function $g(z)$, I find that an industry z displays the home market

[8]Now it becomes more clear why it is helpful to limit the discussion to a sample of country pairs in which the larger country also has higher military expenditure: otherwise the prediction of how production varies with ME/ME^* flips.

effect if and only if $g(z) > 0$.

$$\begin{aligned}
\text{Let } h(z) &= \frac{\tilde{n}(z)}{\tilde{n}(z)^*} \\
\text{We know } (\tilde{n}(z) + \tilde{n}(z)^*)c(z)\sigma(z) &= \alpha(Y + Y^*) \\
\Rightarrow \frac{h(z)}{h(z)+1}(Y + Y^*) &= \frac{\tilde{n}(z)c(z)\sigma(z)}{\alpha} \\
(Y + Y^*)\left[\frac{h(z)}{h(z)+1} - \frac{Y}{Y+Y^*}\right] &= \frac{\tilde{n}(z)c(z)\sigma(z)}{\alpha} - Y \\
&= g(z) \\
\text{So } h(z) = \frac{\tilde{n}(z)}{\tilde{n}^*(z)} > \frac{Y}{Y^*} &\Leftrightarrow g(z) > 0 \qquad (1.7)
\end{aligned}$$

Similarly, the military sector displays the home market effect if and only if $\tilde{n}_m/\tilde{n}_m^* > \frac{ME}{ME^*} \Leftrightarrow g_m > 0$.

To see the intuition behind this result, note that the g function reflects the trade-off between production costs (represented by w^σ) and trade costs (represented by $x = \tau^{\sigma-1}$):

$$\begin{aligned}
g(z) &> 0 \\
\Leftrightarrow \frac{Y}{x(z)w^{\sigma(z)} - 1} &> \frac{Y^*}{x(z)w^{-\sigma(z)} - 1} \qquad (1.8)
\end{aligned}$$

Equation (1.8) shows the tradeoffs faced by a civilian firm that produces a variety of good z as it considers relocating from the small to the large country: the left hand side portrays the benefits of relocation (higher for a larger Home market, smaller with a higher production cost $w^{\sigma(z)}$), while the right hand side shows the costs of relocation. The industry will show home market effects if the benefits of relocation exceed the costs.

High-x and low-σ industries have relatively high g and so are more likely to show home-market effects. Assuming military expenditure is small relative to GDP for all countries, there will be some civilian industries for which $g(z) > 0$ and some for which $g(z) < 0$. The following result from Hanson and Xiang (2004) holds for civilian sectors:

> Let z_0 be a civilian industry so that $g(z_0) > 0$; then $g(z_1) > 0$ for all z_1 such that $x(z_1) \geq x(z_0)$ and $\sigma(z_1) \leq \sigma(z_0)$. Conversely, if $g(z_0) < 0$ for some z_0, then $g(z_1) < 0$ for all z_1 such that $x(z_1) \leq x(z_0)$ and $\sigma(z_1) \geq \sigma(z_0)$.

In other words, if a civilian industry shows home market effects, so will all industries that have at least as high effective trade costs and are at least as differentiated.

In addition to this, $g(z)$ decreases monotonically with σ for all parameter values, and increases with x as long as an additional condition is met.[9]

In comparing the $g()$ functions of two civilian industries z and z_0, the only parameters of interest were the effective trade cost $x(z)$ and elasticity of substitution $\sigma(z)$. However, as I extend this result to the military sector, and compare g_m to $g(z_0)$, the set of parameters extends by the relative ratio of military spending out of GDP - remember the equilibrium condition:

$$0 = Y\int_0^1 \left[\frac{1}{x(z)w^{\sigma(z)}-1} - \frac{Y^*/Y}{x(z)w^{-\sigma(z)}-1}\right]\alpha(z)dz + ME\left[\frac{1}{x_m w^{\sigma_m}-1} - \frac{ME^*/ME}{x_m w^{-\sigma_m}-1}\right]$$

Whether $g_m > 0$ or $g_m < 0$ depends on how transport costs and the elasticity of substitution compare across sectors, but also on military and civilian budgets in the two countries: i.e. if $ME/ME^* \gg Y/Y^*$ the military sector is much more likely to display the home market effect.

> **Proposition 1**
>
> Let z_0 be a civilian industry so that $g(z_0) > 0$; then $g_m > 0$ if $x_m \geq x(z_0)$, $\sigma_m \leq \sigma(z_0)$, and $ME/ME^* \geq Y/Y^*$. In particular, I isolate two cases:
>
> (a) $x_m > x(z_0)$, $\sigma_m < \sigma(z_0)$, and $ME/ME^* \geq Y/Y^*$

[9]The condition is that the two countries are not too different in size, or that effective trade costs are not too high: $Y/Y^* < \frac{(x(z)w^{\sigma(z)}-1)^2}{(x(z)-w^{\sigma(z)})^2}$. But a similar condition is built in implicitly in the assumption of incomplete specialization: if one country were much larger than the other, then at least some high transport cost industries would locate exclusively in the large market.

(b) $x_m \approx x(z_0)$, $\sigma_m \approx \sigma(z_0)$, and $ME/ME^* \geq Y/Y^*$

The reverse also holds: if $g(z_0) < 0$ for some z_0, then $g_m < 0$ if $ME/ME^* \leq Y/Y^*$, $x_m \leq x(z_0)$ and $\sigma_m \geq \sigma(z_0)$.

Proposition 1 states that if a civilian industry z_0 shows home market effects, so will the military industry, as long as the military sector has at least as high effective trade costs and is at least as differentiated as z_0, and as long as Home's military spending relative to Foreign is higher than Home's civilian spending.

As we switch from comparing two civilian industries to comparing military vs. civilian goods, the key difference is that home market effects can arise not just from differences in goods' characteristics, but also from differences in relative demand for military vs. civilian goods (as shown in part b of proposition 1). The military sector is much more likely to display home market effects if Home has higher military spending relative to GDP than Foreign.

1.2.5 Empirical specification

Similarly to the empirical approach from Hanson and Xiang (2004), I construct a double-difference specification, comparing two goods exported by two countries to a common importer.

Let τ_{ijk} be iceberg transport costs for industry i between countries j and k, and assume the following form: $\tau_{ijk} = d_{jk}^{\gamma_i}$, where $\gamma_i > 0$ and d_{jk} is the distance between j and k.

Total sales in industry $i \in \{z, m\}$ by country j to country k are:

$$\begin{aligned} \text{for civilian industries: } S_{zjk} &= \alpha_z Y_k n_{zj} \left(\frac{P_{zjk}}{P_{zk}}\right)^{1-\sigma_z} \\ \text{for military: } S_{mjk} &= ME_k n_{mj} \left(\frac{P_{mjk}}{P_{mk}}\right)^{1-\sigma_m} \end{aligned}$$

where P_{ijk} is the delivered c.i.f. (including cost, insurance, freight) price in country k of a good from industry i produced in country j, and P_{ik} is the CES price index for industry i in country k.

$$\begin{aligned} P_{ijk} &= P_{ij}t_{ijk}(d_{jk})^{\gamma_i} \\ &= \left(\frac{\sigma_i}{\sigma_i - 1}\right) w_{ij}t_{ijk}(d_{jk})^{\gamma_i} \end{aligned}$$

where P_{ij} is the f.o.b. (free-on-board) price of a product in industry i manufactured in country j, t_{ijk} is (1 + ad-valorem tariff in k on imports of i from j), w_{ij} is the unit production cost of i in country j, and d_{jk} is the distance between countries j and k.

Compare country j's exports of good i to country k with some other country h's exports.

$$\begin{aligned} \frac{S_{ijk}}{S_{ihk}} &= \frac{n_{ij}}{n_{ih}} \left(\frac{w_{ij}}{w_{ih}}\right)^{1-\sigma_i} \left(\frac{d_{jk}}{d_{hk}}\right)^{(1-\sigma_i)\gamma_i} \\ &= \frac{\tilde{n}_{ij}}{\tilde{n}_{ih}} \left(\frac{w_{ij}}{w_{ih}}\right)^{-\sigma_i} \left(\frac{d_{jk}}{d_{hk}}\right)^{(1-\sigma_i)\gamma_i} \end{aligned} \tag{1.9}$$

Implicit in equation (1.9) is the assumption that countries j and h face common tariffs in country k.

Note that in this first difference the variables specific to civilian vs. military sectors ($\alpha(z), Y, ME$) have already been eliminated, so next in the double-difference there is no reason we cannot compare the military as a treatment industry with a low-transport cost, high-substitution elasticity civilian sector as the control industry.

Let i be the treatment industry (low substitution elasticity and high transport costs) and o the control industry (high substitution elasticity and low transport costs). Then, applying another difference to equation (1.9), I get:

$$\frac{S_{ijk}/S_{ihk}}{S_{ojk}/S_{ohk}} = \frac{\tilde{n}_{ij}/\tilde{n}_{ih}}{\tilde{n}_{oj}/\tilde{n}_{oh}} \frac{(w_{ij}/w_{ih})^{-\sigma_i}}{(w_{oj}/w_{oh})^{-\sigma_o}} (d_{jk}/d_{hk})^{(1-\sigma_i)\gamma_i - (1-\sigma_o)\gamma_o} \tag{1.10}$$

Double-difference, two civilian industries

Equation (1.10) and result (1.7) suggest the following regression for when both i and o are civilian industries:

$$ln\left(\frac{S_{ijk}/S_{ihk}}{S_{ojk}/S_{ohk}}\right) = \alpha+\beta\ln(Y_j/Y_h)+\phi(X_j-X_h)+\theta ln(d_{jk}/d_{hk})+\varepsilon_{iojkh} \quad (1.11)$$

where Y_j and Y_h represent civilian expenditure in the two countries. Since in my sample military expenditure is on average around 2 percent of GDP, and does not exceed 6 percent, I approximate civilian expenditure Y by GDP: $Y = GDP - ME \approx GDP$. Using GDP will also make results easier to interpret. X_j and X_h control for production costs of industries i and o in the two exporter countries, and d_{jk} and d_{hk} are distances from each of the exporters to the common importer

To see how I obtained equation (1.11), imagine an experiment in which we randomly draw the relative size of the two countries. Then when $Y_j > Y_h$, $\frac{\tilde{n}_{ij}/\tilde{n}_{ih}}{\tilde{n}_{oj}/\tilde{n}_{oh}} > 1$, whereas when $Y_j < Y_h$, $\frac{\tilde{n}_{ij}/\tilde{n}_{ih}}{\tilde{n}_{oj}/\tilde{n}_{oh}} < 1$. In other words, $\ln\left(\frac{\tilde{n}_{ij}/\tilde{n}_{ih}}{\tilde{n}_{oj}/\tilde{n}_{oh}}\right)$ varies positively with $\ln(Y_j/Y_h)$.

Thus, equation (1.11) simplifies to the Hanson and Xiang (2004) specification: we test for home market effects by examining whether bigger countries export relatively more highly differentiated, expensive to ship goods - i.e. $\beta > 0$. This estimation approach uses the fact that the exporter pair is not ordered by size, so that $\ln(Y_j/Y_h)$ can take both positive and negative values.

Double-difference, military vs. civilian industries

If industry i is military and o is a civilian industry of equal or lower transport costs and equal or higher σ, proposition 1 suggests that $\frac{\tilde{n}_{ij}/\tilde{n}_{ih}}{\tilde{n}_{oj}/\tilde{n}_{oh}}$ will be increasing in $\frac{ME_j/ME_h}{Y_j/Y_h}$. That result was obtained under the condition that $Y_j > Y_h$ and $ME_j > ME_h$, therefore I order exporter pairs so that the first exporter (j) is larger, and I restrict the sample so that exporter 1's military expenditure is also larger than that of exporter 2 (h).

I then estimate the regression:

$$ln\left(\frac{S_{ijk}/S_{ihk}}{S_{ojk}/S_{ohk}}\right) = \alpha + \beta \ln\left(\frac{ME_j/ME_h}{Y_j/Y_h}\right) + \phi(X_j - X_h) + \theta ln(d_{jk}/d_{hk}) + \varepsilon_{iojkh} \quad (1.12)$$

where $\frac{ME_j/ME_h}{Y_j/Y_h}$ is the relative military spending out of GDP of the two exporters, and again X_j and X_h control for the production costs of industries i and o in the two exporter countries, and d_{jk} and d_{hk} are distances from each of the exporters to the common importer. A positive β coefficient is evidence of the home market effect.

Civilian goods o are first taken to be *control* goods (lower transport costs and higher elasticity of substitution), then I consider *similar* goods (approximately equal transport costs and differentiation). I expect the β coefficient to be positive and significant in both cases, but higher in the former case.

1.3 Data

1.3.1 Data sources

I use bilateral trade data from UN Comtrade, classified by the Harmonized System at the 6-digit level, from 1988 through 2007.[10]

To proxy for the distance variable d from equation (2.5), I use both physical distance, and distance in terms of cultural and religious similarity: I obtained inter-capital distance data and indicators of common language, contiguity, and past colonial relationship from *Centre d'Etudes Prospectives et d'Informations Internationales* (CEPII). Thomas Baranga kindly provided his index measure of religious similarity (as constructed in Baranga, 2009).

The X vector of variables is intended to control for production costs across industries, and it includes capital per worker data from the Penn World Tables 5.6,

[10] 1988 is when some countries started reporting trade according to the Harmonized System, but this didn't become the standard method of reporting until the early 1990's (the United States report in HS starting in 1991). It's possible there is some selection over the countries that transfer to the Harmonized System faster, but I am using importer reports, and the importer is differenced out in the analysis.

average total years of education from the Barro and Lee (2001) dataset, and land per worker from the World Bank's World Development Indicators (WDI). GDP and military expenditure as a share of GDP were also extracted from the World Bank's WDI.[11]

Data on international conflicts is from the UCDP/PRIO Armed Conflict Dataset, as described in Gleditsch et al. (2002), version 4-2008.

In the section analyzing determinants of military expenditure, I also use Polity score from the Polity IV Project.

1.3.2 Determinants of military expenditure

My empirical strategy relies on the assumption that military expenditure is exogenous to economic considerations which might influence arms and civilian goods exports.

I investigate this assumption using econometric analysis, and find that military spending reflects country idiosyncracies and is very persistent over time: over all 156 countries in the 20-year unbalanced panel, 81.1% of variation in the log[12] of military expenditure as a share of GDP is explained by country fixed effects.

For regression results over the 25 OECD countries, see table (1.1). The left-hand side variable is the log of military expenditure share of GDP, and the sample is restricted to be the same across regressions, for easier comparison. As the first column shows, 85.5% of variation is explained by country fixed effects. Alternatively, 98.6% is explained by the previous year's value.

There is only limited correlation with income level and business cycles: there is no clear correlation over the whole sample between $\ln(ME/GDP)$ and $\ln(GDP)$. For OECD countries, the correlation is positive. After accounting for country and year fixed effects, there is still no correlation for the whole sample, but negative correlation for

[11] The military expenditure data is obtained by the World Bank from the Stockholm International Peace Research Institute (SIPRI).

[12] I consider the natural log of military expenditure and GDP in this analysis, in order to limit the influence of single country outliers like the United States, and because the main analysis is in logs. Results are the same in levels.

OECD countries (this is likely to be merely a coincidence as the sample period catches the end of the Cold War, and GDP trends up).

Form of governance matters: the Polity score is negatively correlated with military spending, even after controlling for country and income level - democracies spend less, autocracies more. In particular, strong constraints on the executive predict lower spending.

Finally, spending is higher during conflicts and neighboring conflicts, however the effect is small or zero after controlling for country fixed effects (compare regressions 4 and 3), suggesting that long-lasting regional tensions influence military spending more than individual outbursts of violence. Nonetheless, conflict involvement will prove to be a helpful tool in my analysis, as I will show that conflicts lead to higher arms exports through their positive effect on military expenditure.

1.3.3 Set of military goods

My goal is to analyze trade in military arms, as separate from other types of weaponry. The Harmonized System has an arms and ammunition category (2-digit code 93), from which I eliminate goods which have mainly civilian or law enforcement uses: sporting firearms, signal pistols; compressed air and spring-operated guns; cartridges used for riveting/captive bolt guns etc., handguns and their parts, as well as shotguns and parts, shotguns cartridges and airgun pellets (see table 1.2).

I've added to the sample tanks and other armored vehicles (HS category 8710). Unfortunately, military planes and ships are not distinguishable from civilian vehicles and cannot be used in the analysis.[13] Note that by excluding these goods of extreme R&D and other initial fixed costs, I am left with a set of goods of more moderate returns to scale.

[13]Warships are identified separately, but only starting with the 2002 revision of the Harmonized System.

1.3.4 Substitution elasticities and freight costs

To test the model predictions, I need information on goods' substitution elasticities and freight costs.

For the substitution elasticity parameter, there are two recent estimates in the trade literature: Hummels (1999) and Broda and Weinstein (2006) estimates. Hanson and Xiang (2004) use Hummels' elasticities, but I am not able to do so, due to aggregation problems.[14] I therefore use Broda and Weinstein (2006) estimates for the elasticity of substitution. Hanson and Xiang (2004) estimate freight rates by using US imports data, and employ use their results here.

The full arms and ammunition category (classified under sector 891 in SITC revision 3, and sectors 93 and 8710 in the Harmonized System) is characterized by high differentiation, but relatively low transport costs: it falls within decile 2 of both σ and the freight rate (see figure 1.1). In order to complete the empirical exercise, I select two sets of goods: a group of *similar* industries and a group of *control* goods, as detailed below.

Similar industries

I isolate a set of industries that display similar freight rate and substitution elasticity to the arms and ammunition category. Specifically, I impose that similar goods lie within a 30% radius of weapons, as measured in the Euclidean (σ-1;freight rate) space - see figure (1.2) for the parameter placement of identified similar goods. I also require that these goods have a straightforward correspondence into the Harmonized System classification, since the trade data I use are under this classification. Table (1.4) lists the 7 similar industries selected.

[14] Hummels calculates σ at the Standard International Trade Classification (SITC) revision 3 2-digit level only, which is unusable here, since arms and ammunition are category 891, and at the 2-digit level they are grouped with printed matter, toys etc. in category 89 - "miscellaneous manufactures".

Control industries

The main empirical comparison is done with so-called *control* goods - goods that are cheaper to ship and less differentiated than military weapons, and so are expected to display lower home market effects. I isolate a set of 16 industries that can be used as controls - these have lower freight rate and higher substitution elasticity σ than arms and ammunition, but lie outside the 30% radius populated by similar goods (see the preceding paragraph). Figure (1.3) demonstrates how control goods lie in the South-East quadrant relative to arms in the (σ, freight rate) coordinates. Again, I require that these industries have an easy correspondence between their SITC rev. 3 categorization and the Harmonized System. From the industries listed in table (2.4), I eliminate aircraft (since they include military planes) and engines (this category includes jet engines, which are used for military jet fighters, guided missiles and unmanned aerial vehicles).

1.3.5 Empirical strategy

I estimate equation (2.5) for military vs. *control* industries and interpret the result of $\beta > 0$ as evidence of the home market effect, as resulting from either differences in goods' characteristics or differences in demand. Secondarily, I estimate equation (2.5) again, this time for arms vs. *similar* goods: $\beta > 0$ now shows the home market effect as stemming from differences in demand alone, therefore I expect this coefficient to be smaller.

1.4 Empirical tests

I run a double-difference specification, where the first difference is arms vs. control goods, and the second is exporter 1 vs. exporter 2. In all specifications I require that exporter pairs have the same NATO membership status, in order to limit the role played by strategic interests. Importer-specific tariffs or regulations are differenced out. Years when either of the exporters was engaged in a conflict are included in the estimation sample and controlled for, but results are robust to excluding conflict periods. I also control for exporters' capital, land and education levels, as well as dyadic exporter-importer relationships, including common language, common border, distance, and having ever been in a colonial relationships.

The variable *conflict* is the difference between conflict involvement indicators for the two exporters, so it can take three values: 1 if exporter 1 is in an international conflict and exporter 2 is not, 0 if either both or none of the exporters are at war, and -1 if only the second exporter is involved in conflict. *Colonial relationship*, *common border* and *common language* are also simple differences, this time of dyadic dummy variables relating the two exporters to the importer. *Distance* is the log difference (log of the ratio) of distances between each of the two exporters and the importer. *Religious similarity* is a simple difference between the index of religious closeness between exporter 1 and importer vs. exporter 2 and importer. Since the original index takes values between 0 and 1, the difference ranges between -1 and 1. Finally, *capital per worker*, *land per worker*, and *years of schooling* are log differences between endowment levels of exporters 1 and 2. Table (1.5) shows summary statistics for all included variables over the baseline estimation samples.

1.4.1 Baseline results and exporter sample robustness

I explore three groups of exporters: the top 60[15] economies, a smaller set of just the top 30 countries, and finally a sample of high income OECD countries.[16]

The set of importers is made up of the top 60 economies by GDP each year.[17] Consistent with the theoretical derivation, exporters are ordered so that exporter 1 is larger, and the sample is trimmed so exporter 1 also has higher military expenditure.[18]

[15]The top 60, and for the second sample the top 30, are chosen by GDP in 1992.

[16]I use the World Bank classification based on 2006 GNI.

[17]I select importers by size during the year of trade, in order to capture the largest countries with the most trade, whereas when selecting exporters, I choose a stable sample across years, in order to effectively use lags as instruments in some specifications. If I select the set of importers using the same criterion (by 1992 GDP), results are the same but the sample is reduced.

[18]Trimming the sample so that the larger exporter also has higher military expenditure reduces the top

I include year and importer dummies, and use clustered standard errors to control for correlation within an exporter pair's observations.

In the estimation of equation (2.5), a positive and significant coefficient on the log-differenced ratio of military expenditure to GDP is interpreted as evidence of the home market effect. Results in table (1.6) show coefficients not just above zero, but above 1, which implies a large home market effect: a 10% increase in military expenditure is associated with a 12-16% increase in arms exports, so the impact on domestic production must be much higher than one-for-one.

For all three samples of exporters, I conducted robustness checks over the set of control variables included, as well as the sample of exporters. Columns 1, 3, and 5 in table (1.6) show results without any controls, which maximizes the sample size. Note that the coefficient of interest is virtually unchanged when we introduce controls and the sample of exporters is reduced due to missing data. Three exporters drop completely from the high income OECD sample: the Czech Republic, Luxembourg and Portugal. From the top 60 and top 30 samples, I lose several countries as well, including three of the largest arms exporters: Brazil, China, and Russia. There is concern of potential selection bias: perhaps the type of countries with more complete data are systematically different from the countries that are excluded from the sample, in a way which alters results. In regressions not reported here, I test for this by estimating the same limited regression (excluding factor endowments) over the full sample, and then again on the restricted sample. By restricting the sample of exporters the relevant coefficient stays constant, and even appears to drop in the top 30 sample. This suggests there is no bias, or possibly an attenuation bias, from the sample selection due to missing data.

Another possible concern is that results merely capture a systematic effect related to country size - perhaps large countries tend to have relatively higher military spending, and also export more weapons. However, including the log difference of GDP's does not significantly alter the coefficient on military expenditure. The

60 sample by 12 percent, the top 30 sample by 9 percent and the OECD sample by 6 percent. Results are similar if the trimming and/or ordering are not performed.

coefficient on $\ln(GDP)$ is significantly negative for the top 60 and OECD samples, suggesting that, all else equal, larger countries export fewer arms.

The coefficients aside from military expenditure are also interesting to interpret: *conflict* is positive across the three samples, although it is only estimated precisely for the large sample (top 60 economies). Recall that the excess capacity argument is that countries with high military spending may intentionally over-produce and export excess weapons during peacetime, in order to maintain production capacity and have guaranteed means of defending themselves in the event of a conflict. This mechanism would lead to the observed pattern that high military spenders are also the highest arms exporters, even in the absence of economies of scale. However, if this argument holds, we would also expect arms exports to drop significantly during conflicts, in other words the coefficient on *conflict* in table (1.6) would be negative. Even if the excess capacity mechanism was weaker than is necessary to drive the main result, we'd still expect that conflicts negatively impact arms exports, which is why the positive or zero coefficients I obtain are unexpected. For the top 60 sample, the coefficient of .63 is significant not only statistically, but also economically: it suggests that, all else equal, a country in conflict exports 63 percent more arms relative to civilian goods than a country at peace.

The coefficient on *distance* is insignificant. This makes sense once we remember that the freight rate for arms and ammunition was low: only in the second decile as compared to other goods. Although control civilian goods were chosen to have even lower transport costs, the difference between the two types of goods is still small enough not to lead to significant differences in trade based on distance to importer. Having been in a colonial relationship with the importer predicts lower arms exports, although the result is only significant for the top 60 sample. Common language and common border are positive, which is consistent with expectations that good communication and proximity are more important when it comes to trading arms than in the trade of homogeneous civilian goods. Religious similarity appears to be very important for the arms trade: an exporter that shares the same unique religion with an importer will ship between 54 and 133 percent more weapons relative to civilian goods than an exporter which has no religion in common with the importer. Capital and land per worker both have positive impacts on arms exports. While the result for capital is intuitive, since weapons manufacturing is a capital-intensive industry, it's not as obvious why land per worker is so important. One possibility is that having long borders to defend has forced geographically large countries to gain proficiency in manufacturing weapons. Or perhaps countries which have historically been good at producing arms were better able to maintain the integrity of their physical borders. And finally, education has a weakly negative impact on arms exports.

In order to dispel concerns that military aid may be inducing the pattern observed, I try excluding the top four arms exporters[19] from the data, first one by one and then cumulatively, for all three samples, and find results to be robust. Since the top arms exporters are also the top military aid donors, with remaining countries not involved in sizeable aid programmes to my knowledge, this result shows that the pattern I obtain is not due to the mechanical link that military aid establishes between military expenditure and arms exports. I also demonstrate this way that no single large arms exporter is responsible for driving results.

Table (1.15) reports regression outcomes when dropping top exporters one-at-a-time for the top 30 economies.[20]

1.4.2 Instrumental variables estimation

In section 1.3.2 I argued that military expenditure as a share of GDP is persistent across time, and not linked to the type of economic factors that might influence exports. In order to dispel any lingering concerns about the exogeneity of military spending, I run instrumental variable regressions, using two separate instruments:

The first IV strategy is to employ own lags to instrument for the current value

[19] The top arms exporters in my estimation sample are the United States, Germany, France, and the United Kingdom. The Russian Federation and China drop out of the sample when I include endowment controls.

[20] I report results for the top 30 sample, because it has the lowest point estimate of the coefficient on military expenditure.

of the log differenced ME/GDP ratio. The intuition is that, while it is conceivable that changing economic or geopolitical factors may influence both exports and military spending, they cannot change past values of military expenditure. Therefore lags are valid instruments for military spending. Results are reported in table (1.8) and are very robust for five and ten-year lags, despite the shrinking sample size (see the top two rows of table 1.7 for first stage results).

The second instrument I use are contemporaneous or lagged conflicts. By isolating the component of military expenditure due to conflict involvement, I expect the relevant coefficient (on $\ln(ME/GDP)$) to be negative if the excess capacity argument is true. What I find instead is that results are similar to the OLS estimation, although less precisely estimated. *Conflict* is included in the baseline OLS specification, so naturally I don't expect it to be a valid instrument. The point of this exercise is to show that the direction of the bias it introduces is different from the one we might have expected: not only do conflicts not reduce arms exports, but the increase in military expenditure due to conflicts leads to higher exports. See table (1.7) for first stage, and table (1.9) for second stage results.

Since IV results were shown to be no weaker than OLS, I will use ordinary least squares estimation going forward, for the sake of simplicity.

1.4.3 Product sub-samples

Now that we've found evidence of the home market effect for the full category of military arms and ammunition (excluding aircraft and watercraft), the next question is: do results hold for sub-samples as well? Isolating individual 6-digit arms categories leads to loss of estimation precision in most cases, since the regression sample is severely reduced. There is, however, a useful grouping of products we can exploit here: by degree of differentiation. I expect the home market effect to be stronger for goods that are more differentiated, and lower for more homogeneous goods.

Table (1.10) shows in italics the sub-categories which I expect to be highly differentiated (also marked by subscript d), while in bold font are cartridges - expected to be relatively homogeneous (also marked by subscript h). In separate regressions, I found that a 10% increase in military spending is associated with a 17-18% increase in exports of highly differentiated military goods, and only 10-11% increase in exports of cartridges. Nonetheless, since any coefficient above zero would be sufficient evidence for the home market effect, the result for cartridges is still sizeable.

In order to tell whether the coefficients on $\ln(ME/GDP)$ are significantly different, I append the two samples and run a pooled regression, interacting each variable with an indicator variable for the high returns to scale group. Table (1.11) shows that, indeed, the difference in the strength of the home market effect for the two groups is statistically significant at the 1% level.

1.4.4 Comparison to *similar* goods

Table (1.12) shows regression results when I compare arms to *similar* goods. By eliminating the variation in goods' characteristics between military and the comparison civilian goods, I isolate the home market effect as stemming from heterogeneity of demand only, as shown in part (b) of Proposition 1. I expect the coefficients on $\ln(ME/GDP)$ to be smaller than in table (1.6) where we used *control* goods for comparison, and indeed, for the first and third samples (top 60 and OECD countries), the relevant coefficient is now smaller.[21]

Interestingly, the coefficient on *conflict* is now better estimated and gains significance in the smaller samples as well. Once again, a positive sign on the conflict variable indicates that countries export more arms during conflicts, even beyond the effect through military expenditure increases.

All other coefficients are similar to those in table (1.6). For the largest sample - top 60 - a pooled estimation over the control and similar goods, interacting all variables with an indicator for control goods, reveals that the only significantly different

[21] Using a pooled regression, I verified that the coefficient on military spending is significantly smaller in samples Top 60 and OECD when I use similar goods for comparison, as opposed to control goods.

coefficient between the two samples is that on the log differenced ratio of military expenditure and GDP.

1.4.5 Institutional quality and similarity

Anderson and Marcouiller (2002) argue that omitting controls of institutional quality biases coefficients in gravity models. I introduced a simple difference in exporters' Polity score in the baseline estimation, and found positive, but insignificant coefficients. The coefficient of interest, on the log differenced ratio of military expenditure and GDP, remains unchanged.

Another concern is that countries may export arms preferentially to political allies or countries that are similar to them in terms of institutional make-up. I considered two additional controls to address this potential source of bias: an index of UN General Assembly voting similarity (see Gartzke and Jo, 2006), and a measure of Polity score distance between each of the exporters and the importer. Again, the home market effect result persists, with no significant changes in magnitude.

1.5 Conclusion

I show that increases in military spending can help countries become successful arms exporters. From another point of view, I demonstrate how seemingly innocuous investments in own arsenals can contribute to international arms proliferation *from* the spending country.

I compare arms exports to exports of homogeneous civilian goods (as well as civilian goods of similar differentiation and transport costs to military goods) and find that countries with large military expenditure relative to GDP export more arms. The magnitude of the effect is high: a 10 percent increase in military spending leads to a 12 to 16 percent increase in exports of arms and ammunition.

This mechanism that associates higher domestic demand with higher exports is known as the home market effect in the international trade literature, and my approach to estimating it is novel in that I use military expenditure to introduce variation in demand within countries, rather than assume identical preferences and use differences in country size to infer trade patterns.

I instrument for military expenditure with its lagged values from up to 10 years before, and find that results are unchanged. I also use conflicts as instrumental variables and find that, perhaps counter-intuitively, even the increase in military expenditure introduced by international clashes leads to higher arms exports.

The pattern holds when I focus on sub-samples of military goods and, consistent with theoretical predictions, the home market effect is stronger for the types of goods we expect to be more differentiated. My results strengthen the case behind *economic geography* trade models - those combining increasing returns to scale and positive transport costs, to the detriment of comparative advantage-based models.

On the public policy side, this paper demonstrates an avenue through which increased military expenditure may have a stimulating effect on the economy, but the current analysis is insufficient to draw welfare conclusions or dictate policy.

In the realm of international relations, my results suggest a simple approach to containing international weapons proliferation: rather than establish complicated trade restriction rules, major arms exporters need only agree to reduce their defense budgets, and their arms exports will naturally drop.

An interesting extension would be to analyze the role played by the government in other industries where central spending accounts for a large share of the buyer market - for instance, the construction industry. In a separate paper, I examine whether investments in transport infrastructure lead to exports of steel, cement, and other construction materials.

Acknowledgements

I thank Gordon Hanson for invaluable guidance throughout this project. I also thank Sarada, Thomas Baranga, Eli Berman, Lawrence Broz, Tai Ming Cheung, Tiffany Chou, Daniel Egel, Roger Gordon, Youjin Hahn, Craig McIntosh, Marc Muendler, Karthik Muralidharan, Philip Neary, Paul Niehaus, James Rauch, Hee Seung Yang, Li Zhou, and seminar participants at the UCSD Economics Department and IR/PS.

1.6 Figures

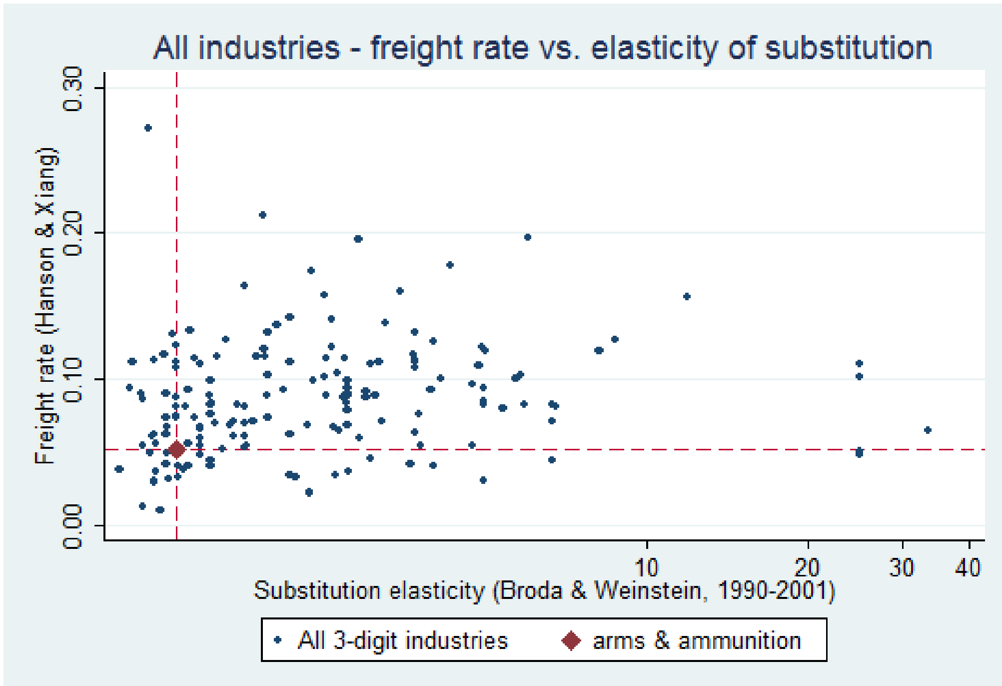

Figure 1.1: Substitution elasticity and freight rate of arms vs. other industries

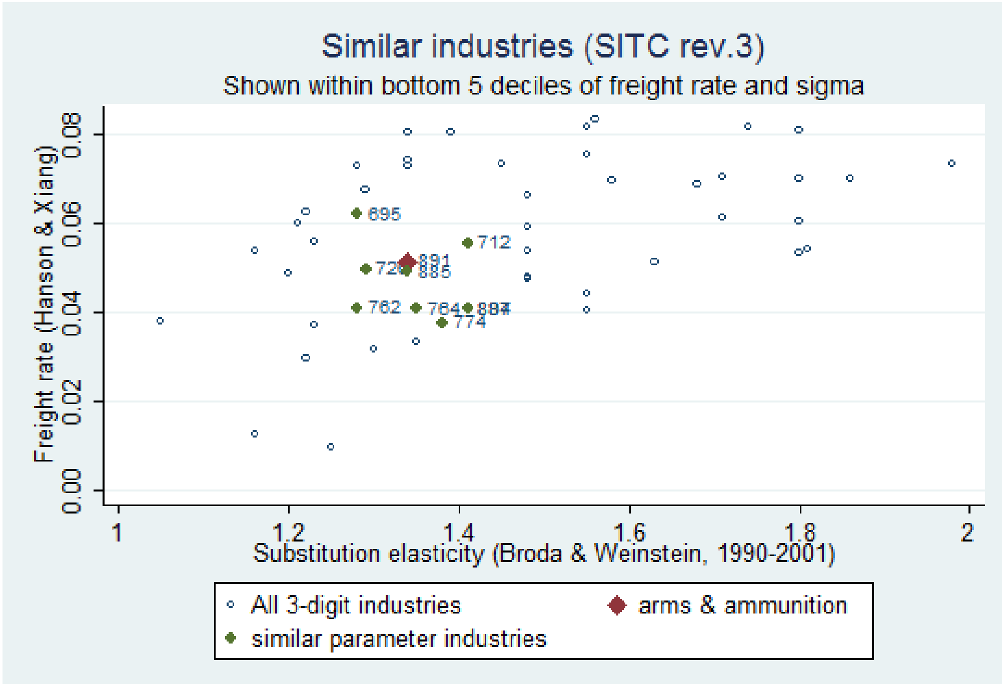

Figure 1.2: Similar goods

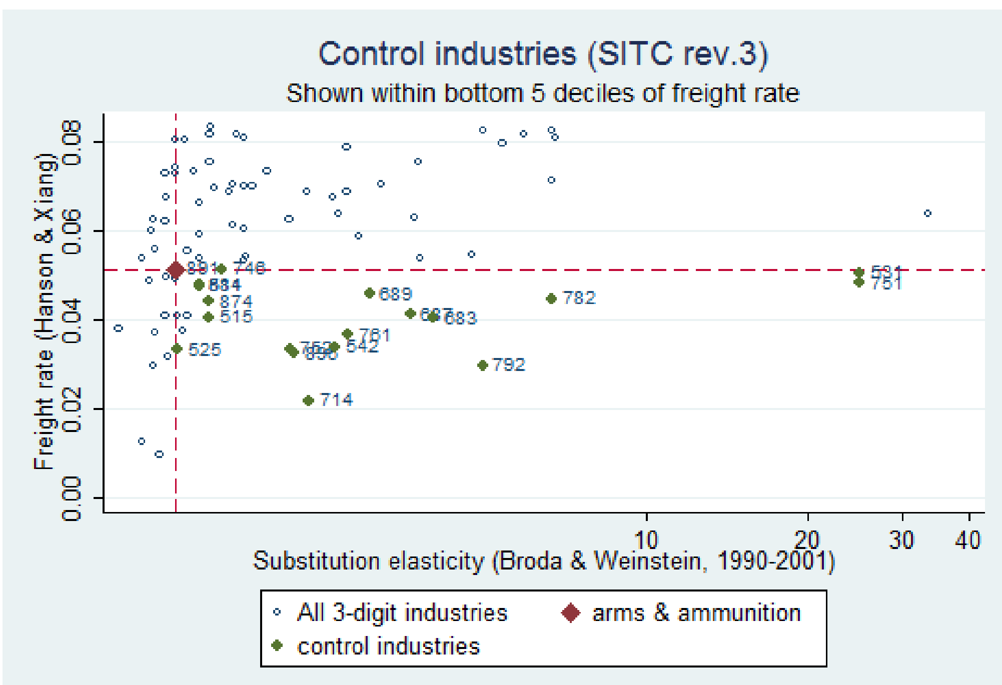

Figure 1.3: Control goods

1.7 Tables

Table 1.1: Explaining variation in $\ln(ME/GDP)$ for OECD countries, years 1988-2007

	s1	s2	s3	s4	s5	s6
	(1)	(2)	(3)	(4)	(5)	(6)
Lag of ln(ME/GDP)					.85	.99
					(.03)***	(.006)***
ln(GDP)			-.49	.05	-.11	
			(.07)***	(.01)***	(.04)***	
Polity score			.06	-.19	.03	
			(.02)***	(.03)***	(.009)***	
Conflict			.10	.57	.05	
			(.03)***	(.10)***	(.02)***	
Neighboring conflicts			-.03	.35	.01	
			(.03)	(.06)***	(.01)	
Country fixed effects	Yes	Yes	Yes		Yes	
Year fixed effects		Yes	Yes		Yes	
Obs.	431	431	431	431	431	431
R^2	.85	.95	.96	.23	.99	.99

Note: The sample is restricted to be the same across regressions.

Table 1.2: Arms and ammunition subcategories eliminated and retained

HS category		Description
8710		Tanks and other armored fighting vehicles
93		Arms and ammunition, parts and accessories thereof
9301		Military weapons, other than hand guns, swords, etc - military rifles, shotguns and other weapons, including: self propelled guns, howitzers, mortar, machine guns, missile and rocket launchers and similar projectors
~~9302~~		~~Revolvers and pistols~~
~~9303~~		~~Other firearms, sporting, etc, signal pistols, etc~~
~~9304~~		~~Arms nes (spring, air or gas guns, truncheons, etc)~~
9305		Parts and accessories of weapons (9301 to 9304)
	~~930510~~	~~Parts and accessories of revolvers or pistols~~
	~~930521~~	~~Shotgun barrels (of heading 9303)~~
	~~930529~~	~~Parts and accessories of shotguns or rifles, nes (of heading 9303)~~
	930590	Parts and accessories nes of weapons, nes (later split into 930591+930599)
9306		Bombs, grenades, mines, missiles, ammunition, etc
	~~930610~~	~~Cartridges for rivet etc tools, humane killers, etc~~
	~~930621~~	~~Cartridges, shotgun~~
	~~930629~~	~~Air gun pellets, parts of shotgun cartridges~~
	930630	Cartridges nes, parts thereof - cartridges for rifle and pistol, empty cartridge shells
	930690	Munitions of war, ammunition/projectiles and parts - bombs, grenades, torpedoes, mines, missiles and sim. munitions of war and parts thereof; other ammunition and projectiles and parts thereof, incl. shot and cartridge wads
9307		Swords, cutlasses, bayonets, lances, scabbards, etc

Note: Categories crossed out have been eliminated from the sample because they have primarily civilian or law enforcement use.

Table 1.3: Set of control goods - less differentiated, cheaper to transport than arms

	SITC rev 3 description	freight rate	σ	HS concordance
714	Engines and motors, nonelectric	0.0217	2.37	8411
792	Aircraft and associated equipment; spacecraft	0.0295	4.98	8801-3, 8805
896	Works of art, collectors' pieces and antiques	0.0323	2.23	9701-6
525	Radioactive and associated materials	0.0331	1.35	2844-6
752	Computers	0.0333	2.18	8471
542	Medicaments	0.0338	2.65	3003, 3004
761	TV receivers	0.0364	2.8	8528
683	Nickel	0.0402	4.04	7502, 7504-7
515	Organo-inorganic compounds	0.0404	1.55	2930-2935
687	Tin	0.0409	3.65	8001, 8003-6
874	Measuring and analysing instruments	0.0440	1.55	9014-7,9023-7,9030-3
782	Motor vehicles for the transport of goods	0.0445	6.7	8704-5
514	Nitrogen-function compounds	0.0475	1.48	2921-9
881	Photographic apparatus and equipment, n.e.s.	0.0477	1.48	9006-8, 9010
531	Synthetic organic coloring matter	0.0504	25.03	3204-5
746	Ball or roller bearings	0.0512	1.63	8482

Table 1.4: Set of similar goods - equal substitution elasticity and freight rate to arms

	SITC rev 3 description	freight rate	σ	HS concordance
897	Jewelry, goldsmiths' and silversmiths' wares	.0405	1.41	7113-7
884	Optical goods, n.e.s.	.0405	1.41	9001-4
764	Telecommunications equipment n.e.s. and parts	.0407	1.35	8517-8, 8522, 8525-6, 8529
762	Radio-broadcast receivers	.0408	1.28	8527
885	Watches and clocks	.0490	1.34	91xx
712	Steam turbines, other vapor turbines and parts	.0553	1.41	8406
695	Tools for use in the hand or in machines	.0620	1.28	8201-9

Table 1.5: Summary statistics for the baseline estimation samples

	Top 60		Top 30		High income OECD	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
dependent variable	-0.12	3.65	-0.08	3.43	-0.04	3.44
ln(ME/GDP)	0.17	0.62	0.23	0.53	0.24	0.49
conflict	0.03	0.52	0.09	0.48	0.10	0.33
distance	0.23	1.15	0.30	1.13	0.26	1.12
colonial relationship	0.01	0.34	0.02	0.35	0.03	0.34
common language	0.01	0.46	0.04	0.44	0.02	0.45
common border	0.00	0.38	-0.01	0.40	0.00	0.39
religious similarity	-0.06	0.26	-0.06	0.26	-0.06	0.25
capital/worker	0.06	1.10	-0.06	0.95	0.00	0.45
land/worker	-0.34	2.03	-0.28	2.11	-0.21	2.13
years schooling	0.07	0.37	0.05	0.35	0.09	0.28
N obs.	37,604		26,877		27,842	

Table 1.6: Baseline estimation

	Top 60		Top 30		High income OECD	
	(1)	(2)	(3)	(4)	(5)	(6)
ln(ME/GDP)	1.51 (.13)***	1.45 (.15)***	1.21 (.21)***	1.19 (.24)***	1.43 (.26)***	1.62 (.28)***
conflict		.63 (.15)***		.15 (.18)		.26 (.19)
distance		-.06 (.07)		.03 (.09)		-.008 (.09)
colonial relationship		-.34 (.16)**		-.12 (.17)		-.08 (.20)
common language		.19 (.13)		.41 (.15)***		.01 (.16)
common border		.06 (.13)		.38 (.13)***		.34 (.13)***
religious similarity		1.33 (.23)***		.93 (.25)***		.54 (.22)**
capital/worker		.45 (.12)***		.28 (.16)*		1.15 (.34)***
land/worker		.39 (.06)***		.35 (.07)***		.33 (.07)***
years schooling		.02 (.31)		-.21 (.44)		-.43 (.49)
Obs.	66,218	37,604	34,647	26,877	31,618	27,842
e(N-clust)	923	312	213	148	155	120
R^2	.08	.16	.06	.12	.07	.14

Notes: dep. variable = $\ln\left(\frac{S_{mjk}/S_{mhk}}{S_{ojk}/S_{ohk}}\right)$: flow of military goods (m) from exporters j and h to importer k, vs. flows of control goods (o). Exporter pairs must have the same NATO status. They are ordered so that exporter 1 has higher GDP, and the sample is restricted so that exporter 1 also has higher military expenditure. ME is the ratio in military expenditure between exporter country pairs, while GDP is the gross domestic product ratio. Year and importer country dummies are included in all regressions. Exporter-pair clustered standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***).

The top 60 and top 30 samples contain the largest economies by GDP in 1992. The high income OECD sample is chosen according to the World Bank classification based on 2006 GNI.

Table 1.7: First stage IV results for tables (1.8) and (1.9)

	Top 60	Top 30	High income OECD
ln(ME/GDP), lag 5	.94 (.01)***	.91 (.02)***	.94 (.02)***
ln(ME/GDP), lag 10	.93 (.02)***	.90 (.04)***	.90 (.04)***
conflict, current	.44 (.11)***	.33 (.07)***	.30 (.07)***
conflict, lag 2	.44 (.12)***	.31 (.06)***	.28 (.07)***

Notes: dep. variable = $\ln\left(\frac{ME_j/ME_h}{GDP_j/GDP_h}\right)$. Standard errors are clustered by exporter pair. Significance indicated is at 10%(*), 5%(**), and 1%(***).

Table 1.8: Instrumental variables estimation, using lags of $\ln(ME/GDP)$ as instruments

	Top 60		Top 30		High income OECD	
Lags as IV:	t-5	t-10	t-5	t-10	t-5	t-10
	(1)	(2)	(3)	(4)	(5)	(6)
ln(ME/GDP)	1.48	1.46	1.15	1.14	1.51	1.54
	(.18)***	(.20)***	(.29)***	(.34)***	(.33)***	(.41)***
conflict	.43	.30	.18	.09	.05	-.06
	(.18)**	(.24)	(.21)	(.27)	(.24)	(.33)
distance	-.07	-.08	.06	.02	.03	-.02
	(.08)	(.08)	(.10)	(.11)	(.08)	(.09)
colonial relationship	-.44	-.21	-.35	-.24	-.19	.13
	(.20)**	(.25)	(.22)	(.28)	(.23)	(.29)
common language	.34	.27	.68	.70	.15	.14
	(.16)**	(.18)	(.18)***	(.21)***	(.19)	(.22)
common border	-.10	.02	.18	.20	.08	.12
	(.14)	(.17)	(.15)	(.19)	(.15)	(.18)
religious similarity	1.59	1.59	1.46	1.63	1.16	1.29
	(.27)***	(.30)***	(.29)***	(.32)***	(.27)***	(.32)***
capital/worker	.42	.39	.19	.12	1.14	1.54
	(.14)***	(.17)**	(.20)	(.23)	(.35)***	(.39)***
land/worker	.36	.35	.33	.32	.32	.29
	(.06)***	(.07)***	(.08)***	(.08)***	(.07)***	(.08)***
years schooling	.61	.89	.51	1.03	.54	.82
	(.36)*	(.41)**	(.49)	(.56)*	(.51)	(.56)
Obs.	17,132	9,714	12,841	7,361	13,821	7,589
e(N-clust)	231	214	119	98	108	97
R^2	.21	.22	.18	.21	.18	.21

Notes: dep. variable = $\ln\left(\frac{S_{mjk}/S_{mhk}}{S_{ojk}/S_{ohk}}\right)$: flow of military goods (m) from exporters j and h to importer k, vs. flows of control goods (o). Exporter pairs must have the same NATO status. They are ordered so that exporter 1 has higher GDP, and the sample is restricted so that exporter 1 also has higher military expenditure. ME is the ratio in military expenditure between exporter country pairs, while GDP is the gross domestic product ratio. Year and importer country dummies are included in all regressions. Exporter-pair clustered standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***).

Table 1.9: Instrumental variables estimation, using conflicts as instruments

	Top 60		Top 30		High income OECD	
Lags of *conflict* as IV:	t	t-2	t	t-2	t	t-2
	(1)	(2)	(3)	(4)	(5)	(6)
ln(ME/GDP)	2.89	2.55	1.64	1.45	2.49	1.46
	(.45)***	(.41)***	(.50)***	(.50)***	(.66)***	(.54)***
distance	.02	-.003	.05	.06	.01	.03
	(.07)	(.08)	(.09)	(.10)	(.09)	(.09)
colonial relationship	-.33	-.41	-.13	-.23	-.08	-.14
	(.16)**	(.17)**	(.18)	(.19)	(.21)	(.20)
common language	.30	.34	.43	.49	.13	.03
	(.15)**	(.15)**	(.16)***	(.17)***	(.19)	(.19)
common border	.13	.06	.38	.31	.30	.24
	(.13)	(.14)	(.13)***	(.14)**	(.14)**	(.13)*
religious similarity	1.70	1.47	.97	.98	.53	.77
	(.28)***	(.28)***	(.24)***	(.24)***	(.21)***	(.23)***
capital/worker	.65	.53	.43	.23	1.65	1.08
	(.18)***	(.18)***	(.24)*	(.26)	(.46)***	(.43)**
land/worker	.36	.35	.35	.34	.30	.31
	(.07)***	(.07)***	(.07)***	(.07)***	(.07)***	(.07)***
years schooling	-.64	-.09	-.57	.07	-1.02	.39
	(.41)	(.43)	(.52)	(.53)	(.66)	(.61)
Obs.	37,604	23,328	26,877	17,514	27,842	18,446
e(N-clust)	312	250	148	125	120	109
R^2	.11	.15	.12	.15	.13	.17

Notes: dep. variable = $\ln\left(\frac{S_{mjk}/S_{mhk}}{S_{ojk}/S_{ohk}}\right)$: flow of military goods (m) from exporters j and h to importer k, vs. flows of control goods (o). Exporter pairs must have the same NATO status. They are ordered so that exporter 1 has higher GDP, and the sample is restricted so that exporter 1 also has higher military expenditure. ME is the ratio in military expenditure between exporter country pairs, while GDP is the gross domestic product ratio. Year and importer country dummies are included in all regressions. Exporter-pair clustered standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***).

Table 1.10: Arms and ammunition subcategories included in the baseline

6-digit HS	Category description
871000[d]	*Tanks and other armored fighting vehicles*
*9301*xx[d]	*Military weapons, other than hand guns, swords, etc* *- military rifles, shotguns and other weapons, including: self propelled guns, howitzers, mortar, machine guns, missile and rocket launchers and similar projectors*
93059x	Parts and accessories of weapons other than handguns and sporting guns
930630[h]	**Cartridges nes, parts thereof** **- cartridges for rifle and pistol, empty cartridge shells**
930690[d]	*Munitions of war, ammunition/projectiles and parts* *- bombs, grenades, torpedoes, mines, missiles and sim. munitions of war and parts; other ammunition and projectiles and parts thereof, incl. shot and cartridge wads*
930700	Swords, cutlasses, bayonets, lances, scabbards, etc

Note: Superscript ^{d} and *italics* mark highly differentiated goods, while superscript ^{h} and **bold font** mark goods that are less differentiated (more homogeneous).

Table 1.11: High and low-differentiation goods - pooled regression

	Top 60	Top 30	High income OECD
	(1)	(2)	(3)
ln(ME/GDP)	1.08	.79	1.08
	(.22)***	(.35)**	(.39)***
ln(ME/GDP)*Differentiated	.87	.87	.73
	(.18)***	(.25)***	(.26)***
Obs.	30,005	23,961	26,445
e(N-clust)	295	144	120
R^2	.14	.12	.13

Notes: Coefficients not shown are the right hand side variables in the baseline regression (conflict, distance, colonial relationship, common language, common border, religious similarity, capital per worker, land per worker, years of schooling and dummy variables for importer and year), as well as interaction terms of all variables with an indicator variable for highly differentiated goods.

Table 1.12: Baseline estimation - arms vs. similar goods

	Top 60	Top 30	High income OECD
	(1)	(2)	(3)
ln(ME/GDP)	1.16	1.18	1.23
	(.16)***	(.26)***	(.31)***
conflict	.69	.32	.55
	(.16)***	(.19)*	(.20)***
distance	-.11	.02	-.02
	(.07)*	(.09)	(.09)
colonial relationship	-.23	.002	-.11
	(.17)	(.20)	(.21)
common language	.21	.39	.18
	(.13)	(.15)***	(.16)
common border	-.05	.21	.19
	(.12)	(.13)	(.12)
religious similarity	1.38	1.02	.65
	(.25)***	(.27)***	(.24)***
capital/worker	.39	.35	1.19
	(.12)***	(.17)**	(.32)***
land/worker	.41	.35	.23
	(.06)***	(.07)***	(.07)***
years schooling	.02	-.43	.12
	(.32)	(.44)	(.48)
Obs.	37,613	26,871	27,842
e(N-clust)	313	148	120
R^2	.14	.12	.12

Notes: dep. variable = $\ln\left(\frac{S_{mjk}/S_{mhk}}{S_{ojk}/S_{ohk}}\right)$: flow of military goods (m) from exporters j and h to importer k, vs. flows of *similar* goods (o).

1.8 Appendix

1.8.1 Derivations and proofs

Solving for the trade equilibrium

Each country's income equals the sum of sales revenue from all its civilian and military goods:

$$\begin{aligned} wL = Y + ME &= \left[\int_0^1 n(z)p(z)q(z)dz\right] + n_m p_m q_m \\ 1 = Y^* + ME^* &= \left[\int_0^1 n^*(z)p^*(z)q^*(z)dz\right] + n_m^* p_m^* q_m^* \end{aligned} \tag{1.13}$$

Let $\tilde{n} = nw$, $\tilde{n}^* = n^*w^* = n^*$. Then, replacing $p(z)q(z) = w(z)c(z)\sigma(z)$, $p^*(z)q(z) = c(z)\sigma(z)$, $p(z) = \frac{\sigma(z)}{\sigma(z)-1}w$ and $p^*(z) = \frac{\sigma(z)}{\sigma(z)-1}$, we can re-write the market clearing conditions (1.2) and (1.3) for civilian goods:

$$\begin{aligned} \tilde{n}c\sigma &= \alpha Y \frac{nw^{1-\sigma}}{nw^{1-\sigma} + n^*x^{-1}} + \alpha Y^* \frac{nw^{1-\sigma}}{nw^{1-\sigma} + n^*x} \\ &= \alpha Y \frac{\tilde{n}}{\tilde{n} + \tilde{n}^* w^{\sigma} x^{-1}} + \alpha Y^* \frac{\tilde{n}}{\tilde{n} + \tilde{n}^* w^{\sigma} x} \\ (\tilde{n} + \tilde{n}^*)c\sigma &= \alpha(Y + Y^*) \\ \Rightarrow \tilde{n}(z) &= \frac{Yx(z)^2 - w^{\sigma(z)}(Y + Y^*)x(z) + Y^*}{x(z)^2 - (w^{\sigma(z)} + w^{-\sigma(z)})x(z) + 1} \frac{\alpha(z)}{c(z)\sigma(z)} \end{aligned} \tag{1.14}$$

We do the same for military goods:

$$\begin{aligned}
\tilde{n}_m c_m \sigma_m &= ME \frac{\tilde{n_m}}{\tilde{n_m} + \tilde{n_m}^* w^{\sigma_m} x_m^{-1}} + ME^* \frac{\tilde{n_m}}{\tilde{n_m} + \tilde{n_m}^* w^{\sigma_m} x_m} \\
(\tilde{n}_m + \tilde{n}_m^*) c_m \sigma_m &= ME + ME^* \\
\Rightarrow \tilde{n}_m &= \frac{ME x_m^2 - w^{\sigma_m}(ME + ME^*) x_m + ME^*}{x_m^2 - (w^{\sigma_m} + w^{-\sigma_m}) x_m + 1} \frac{1}{c_m \sigma_m} \qquad (1.15)
\end{aligned}$$

We replace the formulas for $\tilde{n}(z)$ and $\tilde{n}_m$ from (1.14) and (1.15) into equation (1.13):

$$\begin{aligned}
wL = Y + ME &= \left[\int_0^1 n(z) p(z) q(z) dz\right] + n_m p_m q_m \\
&= \left[\int_0^1 \frac{Y x(z)^2 - w^{\sigma(z)}(Y + Y^*) x(z) + Y^*}{x(z)^2 - (w^{\sigma(z)} + w^{-\sigma(z)}) x(z) + 1} \alpha(z) dz\right] + \\
&\quad + \frac{ME x_m^2 - w^{\sigma_m}(ME + ME^*) x_m + ME^*}{x_m^2 - (w^{\sigma_m} + w^{-\sigma_m}) x_m + 1} \\
\Rightarrow 0 &= \int_0^1 \left[\frac{Y x(z)^2 - w^{\sigma(z)}(Y + Y^*) x(z) + Y^*}{x(z)^2 - (w^{\sigma(z)} + w^{-\sigma(z)}) x(z) + 1} - Y\right] \alpha(z) dz + \\
&\quad + \left[\frac{ME x_m^2 - w^{\sigma_m}(ME + ME^*) x_m + ME^*}{x_m^2 - (w^{\sigma_m} + w^{-\sigma_m}) x_m + 1} - ME\right]
\end{aligned}$$

The equilibrium condition is then:

$$\begin{aligned}
0 &= \int_0^1 \alpha(z) g(z) dz + g_m \\
\text{where} \quad g(z) &= \left[\frac{Y}{x(z) w^{\sigma(z)} - 1} - \frac{Y^*}{x(z) w^{-\sigma(z)} - 1}\right] \\
g_m &= \left[\frac{ME}{x_m w^{\sigma_m} - 1} - \frac{ME^*}{x_m w^{-\sigma_m} - 1}\right]
\end{aligned}$$

Existence of a unique solution

For simplicity, I assume $\min[x(z)^{1/\sigma(z)}] < x_m^{1/\sigma_m} < \max[x(z)^{1/\sigma(z)}]$ - in other words, that the military sector is not at an extreme in terms of this measure combining effective trade costs x and elasticity of substitution σ. This is a reasonable assumption in theory, and it is also holds empirically for the freight rate and σ values I use.[22]

I then show that there exists a unique solution to equation 2.2, and that this solution is reached for w in the interval $1 < w < \min[x(z)^{1/\sigma(z)}]$. The intuition for having $w > 1$ in equilibrium is that large country producers have easy access to the larger market, so they incur lower transportation costs. If production costs were also lower here, no producers would wish to locate in the small country.

Existence

Denote the right hand side of equation 2.2 by $R(w)$. Assuming $Y > Y^*$ and $ME > ME^*$, the following conditions are met, ensuring existence of an equilibrium for $1 < w < \min[x(z)^{1/\sigma(z)}]$:

i. $R'(w)$ exists everywhere on the open interval $(1, \min[x(z)^{1/\sigma(z)}])$, and $R'(w) < 0$. This is straightforward to verify.

ii. $R(1) = \int_0^1 \alpha(z)\frac{Y-Y^*}{x(z)-1}dz + \frac{ME-ME^*}{x_m-1} > 0$.

iii. As w rises toward $\min[x(z)^{1/\sigma(z)}]$, R(w) approaches $-\infty$.

Therefore $R(w)$ must intersect the zero axis for a unique w between 1 and $\min[x(z)^{1/\sigma(z)}]$.

Uniqueness

For all positive intervals excluding $1 < w < \min[x(z)^{1/\sigma(z)}]$, I show that equation (2.2) cannot hold.

i. For $w > \max[x(z)^{1/\sigma(z)}]$, it is easy to verify that $R(w) > 0$.

ii. For $\min[x(z)^{1/\sigma(z)}] < w < \max[x(z)^{1/\sigma(z)}]$, $R(w)$ is ill-defined, since $\exists z$ so that $x(z) - w^{\sigma(z)} = 0$.

---

[22] Recall $x^{1/\sigma} = \tau^{1-1/\sigma} = (1 + \text{freight rate})^{1-1/\sigma}$. Using freight rate estimates from Hanson and Xiang (2004) and elasticity estimates from Broda and Weinstein (2006), I verify that indeed the military sector is in the interior of the civilian industries range in terms of $x^{1/\sigma}$.

iii. Recall that $R(1) > 0$ and $R'(w) < 0$. Then as w decreases from 1 and approaches $\max[x(z)^{-1/\sigma(z)}]$, $R(w)$ increases monotonically towards ∞. Therefore $R(w)$ cannot intersect the axis on this interval.

iv. For $\min[x(z)^{-1/\sigma(z)}] < w < \max[x(z)^{-1/\sigma(z)}]$, $R(w)$ is ill-defined, since $\exists z$ so that $x(z)w^{\sigma(z)} - 1 = 0$.

v. $R(0) = -Y - ME < 0$. As w approaches $\min[x(z)^{-1/\sigma(z)}]$ from below, $R(w)$ falls monotonically towards $-\infty$.

Comparative statics

Using the expressions for $g(z)$ and g_m from (2.3) and (2.4), we have the following:

- As ME increases, holding civilian sector spending Y constant, wage w increases. As ME^* increases, w decreases (foreign wage is normalized at 1). These results are consistent with the earlier finding that relative wage increases with country size.

- As ME increases, holding total output $GDP = wL$ constant, civilian expenditure $Y = GDP - ME$ decreases: there is increased demand for workers in the military, but with less disposable income, consumers demand fewer civilian goods, and so worker demand in the civilian sector drops. Whether demand for workers increases or decreases overall depends on the pattern of trade, which in turn depends on transport costs and good differentiation. If transport costs are infinite (Autarky), the wage remains constant as the relative size of the two sectors changes.

  Let $R(ME, w)$ be the right-hand side of equation (2.2).

$$
\begin{aligned}
R(ME,w) &= \int_0^1 \frac{(GDP-ME)\alpha(z)dz}{x(z)w^{\sigma(z)}-1} - \int_0^1 \frac{(1-ME^*)\alpha(z)dz}{x(z)w^{-\sigma(z)}-1} + \frac{ME}{x_m w^{\sigma_m}-1} \\
&\quad - \frac{ME^*}{x_m w^{-\sigma_m}-1}
\end{aligned}
$$

In equilibrium $R(ME, w) = 0$. Add $\Delta ME > 0$:

$$R(ME + \Delta ME, w) \quad = \quad R(ME, w) + \Delta ME \left[\frac{1}{x_m w^{\sigma_m} - 1} - \int_0^1 \frac{\alpha(z) dz}{x(z) w^{\sigma(z)} - 1} \right]$$

The direction of change in w is undefined in general.

If $\frac{1}{x_m w^{\sigma_m} - 1} > \int_0^1 \frac{\alpha(z)dz}{x(z)w^{\sigma(z)}-1}$ in the original equilibrium (the military has relatively *low* effective trade costs x and *low* substitution elasticity σ compared to the average civilian industry), $R(ME, w)$ has to decrease to attain the new equilibrium point, so w has to increase (since $\frac{\partial R(ME,w)}{\partial w} < 0$).

In other words, holding *GDP* constant, wage increases in the large country when preferences shift towards the sector with more differentiated, cheaper to ship goods: production costs in the large country increase in order to make up for the lower distribution costs suffered by large country producers.

- Now hold *ME*, *GDP* and $GDP^* = 1$ constant, and allow ME^* to increase. Since only ME^* and w are allowed to change, use notation $R(ME^*, w)$ for the right-hand side of equation (2.2).

$$R(ME^* + \Delta ME^*, w) = R(ME^*, w) + \Delta ME^* \left[\int_0^1 \frac{\alpha(z)dz}{x(z)w^{-\sigma(z)} - 1} - \frac{1}{x_m w^{-\sigma_m} - 1} \right]$$

Now the condition for w to increase is that the military sector has higher effective trade costs x and lower substitution elasticity σ than the average civilian industry.

From the perspective of the small country, the wage relative to the large country increases as a result of an increase in military spending (holding GDP constant) as long as military goods have relatively *low* transport costs and *high* elasticity of substitution, i.e. the military sector does not display home market effects. This is because military goods are produced more than proportionately in the small country, so an increase in ME^* means a shift from consumption of imports to consumption of domestic goods, and hence a wage increase.

1.8.2 Robustness to excluding top exporters

I try excluding the top four arms exporters[23] from the data, one by one and cumulatively, for both samples.

Results are reported in tables (1.13) through (1.18). These regressions include the full set of controls used in the main body of the paper - I've just suppressed non-essential coefficients to save space.

The purpose of this exercise is to ensure that no single country is driving results, and indeed, the relevant coefficient maintains its size and significance (and actually becomes uniformly higher when I exclude France). In addition, the top military exporters are also by far the top military aid donors. By finding that results are unchanged with their exclusion, I conclude that military aid is not responsible for the empirical pattern I find.

Table 1.13: Top 60 sample, sensitivity to excluding top exporters one-by-one

Excluded:	none (baseline)	US	Germany	France	UK
	(1)	(2)	(3)	(4)	(5)
ln(ME/GDP)	1.45	1.43	1.57	1.53	1.42
	(.15)***	(.15)***	(.14)***	(.15)***	(.15)***
Obs.	37,604	32,723	33,856	33,634	33,503
e(N-clust)	312	299	282	299	300
R^2	.16	.15	.17	.17	.17

Note: Coefficients not shown: conflict, distance, colonial relationship, common language, common border, religious similarity, capital per worker, land per worker, years of schooling and dummy variables for importer and year.

[23] The top arms exporters in my estimation sample are the United States, Germany, France, and the United Kingdom. The Russian Federation and China drop out of the sample when I include endowment controls.

Table 1.14: Top 30 sample, sensitivity to excluding top exporters cumulatively

Excluded:	none (baseline)	US	US Germany	US Germany France	US Germany France UK
	(1)	(2)	(3)	(4)	(5)
ln(ME/GDP)	1.45	1.43	1.52	1.57	1.51
	(.15)***	(.15)***	(.15)***	(.15)***	(.16)***
Obs.	37,604	32,723	29,623	26,242	23,307
e(N-clust)	312	299	270	259	249
R^2	.16	.15	.17	.18	.2

Note: Coefficients not shown: conflict, distance, colonial relationship, common language, common border, religious similarity, capital per worker, land per worker, years of schooling and dummy variables for importer and year.

Table 1.15: Top 30 sample, sensitivity to excluding top exporters one-by-one

Excluded:	none (baseline)	US	Germany	France	UK
	(1)	(2)	(3)	(4)	(5)
ln(ME/GDP)	1.19	1.09	1.47	1.45	1.19
	(.24)***	(.26)***	(.25)***	(.26)***	(.25)***
Obs.	26,877	22,174	23,351	23,080	22,958
e(N-clust)	148	136	127	136	137
R^2	.12	.1	.14	.13	.13

Note: Coefficients not shown: conflict, distance, colonial relationship, common language, common border, religious similarity, capital per worker, land per worker, years of schooling and dummy variables for importer and year.

Table 1.16: Top 30 sample, sensitivity to excluding top exporters cumulatively

Excluded:	none (baseline)	US	US Germany	US Germany France	US Germany France UK
	(1)	(2)	(3)	(4)	(5)
ln(ME/GDP)	1.19	1.09	1.32	1.54	1.38
	(.24)***	(.26)***	(.26)***	(.29)***	(.30)***
Obs.	26,877	22,174	19,296	16,088	13,335
e(N-clust)	148	136	116	106	97
R^2	.12	.1	.12	.13	.15

Note: Coefficients not shown: conflict, distance, colonial relationship, common language, common border, religious similarity, capital per worker, land per worker, years of schooling and dummy variables for importer and year.

Table 1.17: High income OECD sample, sensitivity to excluding top exporters one-by-one

Excluded:	none (baseline)	US	Germany	France	UK
	(1)	(2)	(3)	(4)	(5)
ln(ME/GDP)	1.62	1.67	1.83	2.14	1.68
	(.28)***	(.33)***	(.29)***	(.29)***	(.28)***
Obs.	27,842	23,032	24,213	23,944	23,832
e(N-clust)	120	108	100	108	109
R^2	.14	.12	.15	.16	.15

Note: Coefficients not shown: conflict, distance, colonial relationship, common language, common border, religious similarity, capital per worker, land per worker, years of schooling and dummy variables for importer and year.

Table 1.18: High income OECD sample, sensitivity to excluding top exporters cumulatively

Excluded:	none (baseline)	US	US Germany	US Germany France	US Germany France UK
	(1)	(2)	(3)	(4)	(5)
ln(ME/GDP)	1.62	1.67	1.83	2.34	2.15
	(.28)***	(.33)***	(.33)***	(.36)***	(.38)***
Obs.	27,842	23,032	20,051	16,742	13,898
e(N-clust)	120	108	89	79	70
R^2	.14	.12	.14	.17	.19

Note: Coefficients not shown: conflict, distance, colonial relationship, common language, common border, religious similarity, capital per worker, land per worker, years of schooling and dummy variables for importer and year.

1.9 Bibliography

Anderson, J.E. and D. Marcouiller, "Insecurity and the pattern of trade: an empirical investigation," *Review of Economics and statistics*, 2002, *84* (2), 342–352.

Baranga, T., "Unreported Trade Flows and Gravity Equation Estimation," 2009.

Barro, R.J. and J.W. Lee, "International data on educational attainment: updates and implications," *Oxford Economic Papers*, 2001, *53* (3), 541–563.

Broda, C. and D.E. Weinstein, "Globalization and the Gains from Variety," *quarterly Journal of Economics*, 2006, *121* (2), 541–585.

Chowdhury, A.R., "A causal analysis of defense spending and economic growth," *Journal of Conflict Resolution*, 1991, *35* (1), 80–97.

Davis, D.R., "The home market, trade, and industrial structure," *American Economic Review*, 1998, *88* (5), 1264–1276.

_ and D.E. Weinstein, "Economic geography and regional production structure: an empirical investigation," *European economic review*, 1999, *43* (2), 379–407.

_ and _ , "Market access, economic geography and comparative advantage: an empirical test," *Journal of International Economics*, 2003, *59* (1), 1–23.

Dixit, A.K. and J.E. Stiglitz, "Monopolistic competition and optimum product diversity," *The American Economic Review*, 1977, *67* (3), 297–308.

Feenstra, R.C., *Advanced international trade: theory and evidence*, Princeton Univ Pr, 2004.

Gartzke, E. and D.J. Jo, "The Affinity of Nations Index, 1946-2002," *accessible at http://www. columbia. edu/eg589/datasets*, 2006.

Gleditsch, N.P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand, "Armed conflict 1946-2001: A new dataset," *Journal of Peace Research*, 2002, *39* (5), 615.

Gold, D., "The Changing Economics of the Arms Trade," *Arming the Future: A Defense Industry for the 21st Century*, 1999.

Hanson, G.H. and C. Xiang, "The Home-Market Effect and Bilateral Trade Patterns," *American Economic Review*, 2004, *94* (4), 1108–1129.

Helpman, E. and P.R. Krugman, *Market structure and foreign trade: Increasing returns, imperfect competition, and the international economy*, MIT press, 1985.

Hewitt, D.P., “Military expenditure: econometric testing of economic and political influences,” *IMF, Washington, DC*, 1991.

Holmes, T.J. and J.J. Stevens, “Does home market size matter for the pattern of trade?,” *Journal of International Economics*, 2005, *65* (2), 489–505.

Hummels, D., “Toward a geography of trade costs,” *University of Chicago*, 1999.

Krugman, P., “Scale Economies, Product Differentiation, and the Pattern of Trade,” *American Economic Review*, 1980, *70* (5), 950–959.

Lumpe, L., “A Framework for Limiting the Negative Consequences of Surplus US Arms Production and Trading,” *Arming the Future: A Defense Industry for the 21st Century*, 1999.

Sandler, T. and K. Hartley, “The economics of defense,” *Cambridge Books*, 1995.

Smith, R., “The demand for military expenditure,” *Handbook of defense economics*, 1995, *1*, 69–87.

Chapter 2

Infrastructure Spending and the Trade in Construction Materials

Abstract

I test the existence of the home market effect for construction materials, by using public infrastructure spending as a measure of demand. The home market effect is the prediction that countries with higher demand for a differentiated good will be net exporters of that good. I construct a theoretical model that suggests goods with high transport costs and high differentiation are most likely to display home market effects. I test this prediction empirically for a handful of construction materials that meet the necessary criteria. As expected, I find that the home market effect holds for alloy steel and construction machinery. However, cement and glass display the opposite trade pattern, whereby increased domestic demand leads to reduced exports. I discuss potential explanations for this result.

2.1 Introduction

Public infrastructure spending is an issue of frequent political debate. It is sizeable portion of the economy: even in the United States, which has lower levels of spending than OECD and many developing countries, total public spending for transportation and water infrastructure in 2007 was 2.4 percent of GDP. Infrastructure investments are generally viewed as a good way to increase a country's capital stock and therefore the productivity of all economic agents within it. Spending on public projects is seen as a particularly attractive policy intervention during economic downturns, because of its immediate effect on aggregate demand.

A number of papers (Aschauer, 1989; Munnell, 1990; Ford et al., 1991; Morrison and Schwartz, 1992; Holtz-Eakin and Schwartz, 1995) examine the effect of public infrastructure spending on private sector productivity. However, there is no corresponding literature investigating the effect of public spending on trade patterns, which are of interest in their own right, through their link to production, employment, geo-political interests, and the trade balance. I examine the link established by the *home market effect* (HME) between government infrastructure spending and the trade in construction materials. HME is the prediction of monopolistic competition models that countries with higher demand for a differentiated good will be net exporters of that good.

Government spending on the infrastructure represents a large portion of the market for many construction materials. This allows me to use government infrastructure spending as a measure of demand in order to test the existence of the home market effect for construction materials, considered individually: does higher demand for construction steel, for instance, lead to higher exports of steel? The aim is to examine whether the set of industries that display home market effects is consistent with theory: monopolistic competition models suggest that HME holds for differentiated goods with high transportation costs.

Hanson and Xiang (2004) introduce a model with a continuum of differentiated-product industries, and exploit differences in country size to demonstrate the existence of the home market effect: they show that industries with high transport costs and low substitution elasticities (i.e., more product differentiation) tend to concentrate in the larger country, while industries with low transport costs and high substitution elasticities (i.e., less product differentiation) concentrate in the smaller country.

Hirakawa (2011) expands upon Hanson and Xiang (2004) by introducing a military sector in addition to the continuum of civilian industries, and uses government defense expenditure as a measure of demand for military goods, in order to demonstrate that the home market effect can arise from differences in government preferences over the strength of the military sector.

The current paper extends this framework beyond the military sector, to investigate the existence of the home market effect in the market for construction materials, once again using government preferences as the source of differential demand which drives production and trade patterns.

Few papers in the empirical trade literature test the home market effect for individual industries. Davis and Weinstein (1999) look at how Japanese regional variation in demand influences production of goods within industries, and find that home market effects matter for eight manufacturing sectors (out of nineteen), including “iron and steel”, “electrical machinery” and “transport equipment”, but not including the “stone, clay, and glass”, or “general machinery” categories. Davis and Weinstein (2003) attempt the same exercise for EU trade, but have less precise estimates.

My paper contributes to the empirical trade literature by further exploring the novel home market effect estimation introduced in Hirakawa (2011), as well as by adding to the existing evidence on which industries display home market effects. From a public policy perspective, it is helpful to more fully understand the impact of infrastructure spending, by looking at trade pattern responses. Since public spending is routinely used as a policy intervention tool to stimulate the domestic economy, it becomes particularly important to understand whether part of that induced demand is transferred abroad, or whether instead there is a more than proportional positive effect on domestic production through the home market effect.

2.2 Theoretical model and predictions

I model two types of goods: a continuum of differentiated consumer product industries, whose goods are demanded by individual consumers, and a group of construction materials, demanded only by the government.[1] I will further assume that government demand for materials has a Cobb-Douglas functional form, so a fixed portion of infrastructure spending is allocated to each material; this is in order to keep the model tractable, although I do not expect qualitative predictions to be different even if spending on different materials were price elastic, within reasonable bounds.

[1] In reality, the private sector also demands construction materials, but I abstract away from this for the sake of simplicity. I acknowledge the concern that, if governments seek to spur demand during economic downturns, infrastructure spending will be counter-cyclical and so ignoring private spending on construction materials may bias results. However, in practice, my sample shows no correlation between GDP and infrastructure spending as a share of GDP, after controlling for country fixed effects.

There is a large country and a small country. Each has one factor of production: labor. The large country has a mass $L > 1$ of workers, each earning wage w. The small country's labor endowment and wage are normalized to 1 (so $w^* L^* = 1$). Each country's infrastructure spending budget I (I^*) is extracted from workers' income by lump-sum taxation, so that workers will have after-tax income $Y = wL - I$ ($Y^* = 1 - I^*$) to spend on consumer goods, while governments spend I (I^*) on construction goods.

2.2.1 Consumer goods

Consumer goods are modeled on a continuum, in order to allow for variation in differentiation and transport costs - the two dimensions that will determine which industries display home market effects. In particular, I consider a continuum of monopolistically competitive industries (as introduced by Dixit and Stiglitz, 1977) indexed by $z \in [0,1]$. Consumers derive utility from purchasing many different varieties of a given product. Firms continue to enter until the last firm just breaks even. Since cost structures are identical across firms, in equilibrium all firms have zero profits.

First, I outline the consumers' problem: individuals have Cobb-Douglas preferences over industries, and constant elasticity of substitution (CES) demand over varieties within an industry:

$$U_{\text{consumer}} = \prod_{z \in [0,1]} \left[\left(\sum_{i=1}^{n(z)} q_{zi}^{\frac{\sigma(z)-1}{\sigma(z)}} \right)^{\frac{\sigma(z)}{\sigma(z)-1}} \right]^{\alpha(z)}$$

In the equation above, $\alpha(z)$ is the consumption share of industry z products and $\int_0^1 \alpha(z)dz = 1$; $n(z)$ is the number of product varieties in industry z, $\sigma(z)$ is the elasticity of substitution between varieties (restricted to be larger than one), and q_{zi} is the quantity of variety i in industry z.

Let $\tau(z) > 1$ be the iceberg transport cost incurred in shipping one unit of output from one country to the other, and $x(z) = \tau(z)^{\sigma(z)-1}$ the effective trade cost[2] for industry

[2]As in all monopolistic competition models, transport costs matter more for industries with high

z.

I will assume there is no international specialization at the industry level, meaning each country produces some goods in each industry. The varieties of industry z are symmetric: let $c(z)$ be the fixed labor requirement, and I normalize the variable labor requirement for each variety to one. Then output and price are the same for all varieties: $q_{zi} = q(z)$, $p_{zi} = p(z)$. As a result of the CES demand specification, the price is a constant markup over marginal cost (in this case, wage w):

$$p(z) = \frac{\sigma(z)}{\sigma(z)-1} w \tag{2.1}$$

Since free entry drives profits to zero, output is fixed and revenues are proportional to fixed costs: $\Pi(z) = p(z)q(z) - [c(z)w + qw] = 0$, and we replace the expression for $p(z)$ from equation (2.1) to find:

$$\begin{aligned} q(z) &= c(z)[\sigma(z)-1] \\ p(z)q(z) &= wc(z)\sigma(z) \end{aligned}$$

2.2.2 Construction materials

The government has Cobb-Douglas demand over different types of construction materials ($\gamma(m)$ indicates the share of government spending allocated to each good: $\sum_{m=1}^{M} \gamma(m) = 1$). Even within a certain class of materials (e.g. construction steel), there are many varieties used in construction projects, each clearly differentiated from the others, and with its specific purpose. Therefore at the variety level demand is CES, as in the case of consumer goods.

---

elasticity of substitution. The exact specification of x will become obvious shortly in the model derivation.

$$U_{\text{government}} = \prod_{m=1}^{M}\left[\left(\sum_{i=1}^{n(m)} q_{mi}^{\frac{\sigma(m)-1}{\sigma(m)}}\right)^{\frac{\sigma(m)}{\sigma(m)-1}}\right]^{\gamma(m)}$$

where $m = 1,\ldots,M$ indexes construction materials, of which we assume there is a finite number. The same industry-level algebra from consumer good z carries through for construction material m.

$$\begin{aligned} p(m) &= \frac{\sigma(m)}{\sigma(m)-1}w \\ q(m) &= c(m)[\sigma(m)-1] \\ p(m)q(m) &= wc(m)\sigma(m) \end{aligned}$$

2.2.3 Trade equilibrium

I arrive at the following equilibrium condition:

$$0 = \int_0^1 \alpha(z)g(z)dz + \sum_{m=1}^{M}\gamma(m)g(m) \tag{2.2}$$

$$\text{where} \quad g(z) = \left[\frac{Y}{x(z)w^{\sigma(z)}-1} - \frac{Y^*}{x(z)w^{-\sigma(z)}-1}\right] \tag{2.3}$$

$$g(m) = \left[\frac{I}{x(m)w^{\sigma(m)}-1} - \frac{I^*}{x(m)w^{-\sigma(m)}-1}\right] \tag{2.4}$$

Both $g(z)$ and $g(m)$ are strictly decreasing in w, so equation (2.2) has a unique solution $w > 1$, as long as $\left[(Y-Y^*)\int_0^1 \frac{\alpha(z)dz}{x(z)-1} + (I-I^*)\sum_{m=1}^{M}\frac{\gamma(m)}{x(m)-1}\right] > 0$, a sufficient condition for which is that both the consumer goods and construction materials sectors of the big country are larger than those of the small country.[3]

[3]Thoroughly demonstrating the existence of this result, under a general number M of construction materials, is an involved analytical exercise that will require certain non-trivial continuity assumptions. A way to simplify the proof is to consider each construction material individually - i.e. set M=1, and use the proof from Hirakawa (2011).

It is then easy to show that functions $g(z)$ and $g(m)$ code the trade-offs in the strategic decision over location faced by firms and, they are the key to whether a certain industry displays home market effects or not. Mirroring the arms paper, I then define the home market effect as follows:

Definition - home market effect

Industry z is said to show home market effects if the country with higher demand for z produces a larger share of world z output than its share of world demand for z. In my 2-country world, that translates to:

a) For civilian industries indexed by z: $\frac{n(z)p(z)q(z)}{n^*(z)p^*(z)q(z)} = \frac{n(z)w}{n^*(z)} > \frac{\alpha(z)Y}{\alpha(z)Y^*} = \frac{Y}{Y^*}$.

Define $\tilde{n}(z) = n(z)w$. Then the condition is $\tilde{n}(z)/\tilde{n}^*(z) > \frac{Y}{Y^*}$ or $n(z)/n^*(z) > \frac{Y/w}{Y^*}$, which I can then show is equivalent to $g(z) > 0$.

b) Under the assumption that the larger country (Home) also has higher infrastructure spending ($I > I^*$), the construction materials sector displays the home market effect if and only if $\tilde{n}(m)/\tilde{n}^*(m) > \frac{I}{I^*} \Leftrightarrow n(m)/n^*(m) > \frac{I/w}{I^*}$. This can be shown to be equivalent to $g(m) > 0$.

I then arrive at the following proposition:

Proposition 1

Let z_0 be a civilian industry so that $g(z_0) > 0$; then $g(m) > 0$ if $x(m) \geq x(z_0)$, $\sigma(m) \leq \sigma(z_0)$, and $I/I^* \geq Y/Y^*$. In particular, I isolate two cases:

(a) $x(m) > x(z_0)$, $\sigma(m) < \sigma(z_0)$, and $I/I^* \geq Y/Y^*$

(b) $x(m) \approx x(z_0)$, $\sigma(m) \approx \sigma(z_0)$, and $I/I^* \geq Y/Y^*$

The reverse also holds: if $g(z_0) < 0$ for some z_0, then $g(m) < 0$ if $I/I^* \leq Y/Y^*$, $x(m) \leq x(z_0)$ and $\sigma(m) \geq \sigma(z_0)$.

Proposition 1 states that if a consumer goods industry z_0 shows home market effects, so will the industry for construction material m, as long as m has at least as high effective trade costs and is at least as differentiated as z_0, and as long as

Home's infrastructure spending relative to Foreign is higher than Home's consumer goods spending.

As before, when we switch from comparing two consumer industries to comparing government-demanded goods vs. consumer goods, the key difference is that home market effects can arise not just from differences in goods' characteristics, but also from differences in relative public to private spending (as shown in part b of proposition 1). The construction materials sector is much more likely to display home market effects if Home has higher infrastructure spending relative to GDP than Foreign.

2.2.4 Empirical specification

The derivation proceeds as in Hirakawa (2011).

Compare country j's exports of goods i and o to country k (S_{ijk} and S_{ojk}), with some other country h's exports, also to country k (S_{ihk} and S_{ohk}). If industry i is a construction material and o is a consumer goods industry of equal or lower transport costs and equal or higher σ, proposition 1 suggests that $\frac{\tilde{n}_{ij}/\tilde{n}_{ih}}{\tilde{n}_{oj}/\tilde{n}_{oh}}$ will be increasing in $\frac{I_j/I_h}{Y_j/Y_h}$. That result was obtained under the condition that $Y_j > Y_h$ and $I_j > I_h$, therefore I order exporter pairs so that the first exporter (j) is larger, and I restrict the sample so that exporter 1's infrastructure spending is also larger than that of exporter 2 (h). I then estimate the regression:

$$ln\left(\frac{S_{ijk}/S_{ihk}}{S_{ojk}/S_{ohk}}\right) = \alpha + \beta \ln\left(\frac{I_j/I_h}{Y_j/Y_h}\right) + \phi(X_j - X_h) + \theta ln(d_{jk}/d_{hk}) + \varepsilon_{iojkh} \quad (2.5)$$

where $\frac{I_j/I_h}{Y_j/Y_h}$ is the relative infrastructure spending out of GDP of the two exporters, X_j and X_h control for the production costs of industries i and o in the two exporter countries, and d_{jk} and d_{hk} are distances from each of the exporters to the common importer. A positive β coefficient is evidence of the home market effect.

Consumer goods o are taken to be *control* goods (lower transport costs and higher elasticity of substitution). At a later time, I will consider *similar* goods (approximately equal transport costs and differentiation). I expect the β coefficient to be positive and significant in both cases, but higher in the former case.

2.3 Data

Internationally-comparable infrastructure spending data are limited. The only international panel data available is the one collected by Eurostat - the statistical office of the European Union - in their *General Government Expenditure by Function database*. Table (2.1) shows the level of disaggregation available in government spending; in red italics I marked the categories most likely to make wide use of construction materials.

As these data are collected and disseminated by Eurostat, and reporting is voluntary, the sample is limited to European countries, in particular a subset of EU member countries. Data availability start years by country are shown in table (2.2); the end year is 2008, which is not constraining, since the trade data I employ runs through 2007.

Table (2.3) lists the expenditure categories initially considered, and the construction materials for which these government expenditures capture a large portion of the market. Note for instance that, while cement is certainly used in housing developments as well, it would be inappropriate to use public spending on housing as a measure of demand for cement, since overall transportation spending is much higher than public housing spending, and different types of public spending are not independent.

I use bilateral trade data from UN Comtrade, classified by the Harmonized System at the 6-digit level, from 1990 through 2007.

To proxy for the distance variable d from equation (2.5), I use both physical distance, and distance in terms of cultural similarity: I obtained inter-capital distance data and indicators of common language, contiguity, and past colonial relationship from *Centre d'Etudes Prospectives et d'Informations Internationales* (CEPII).

The X vector of variables is intended to control for production costs across industries, and it includes capital per worker data from the Penn World Tables 5.6, average total years of education from the Barro and Lee (2001) dataset, and land per worker from the World Bank's World Development Indicators (WDI). GDP was also extracted from the World Bank's WDI.

In this preliminary version of the paper, the set of *control* goods is the one identified relative to the arms sector in Hirakawa (2011), and listed in table (2.4); in the next step I will select a set of control goods separately for each *treatment* good considered (i.e. with higher elasticity of substitution and lower transport costs than the given treatment good).

2.4 Government transportation spending

Government expenditure on transport (which includes building and maintenance of roads, bridges, airports, railways, etc.) is a large spending category, which routinely averages over 2 percent of GDP in OECD countries. There are a handful of easily identifiable large inputs into transport infrastructure construction, including steel, cement, asphalt, and construction machinery.

Graph (2.1) shows the variation of transport expenditure out of GDP vs. GDP for the countries in the sample. It shows that infrastructure spending is relatively stable, but it does fluctuate, and it shoots up during certain periods - this is not the same calendar year across countries, although average spending was higher earlier in the sample (early 1990s) and evened off during the 2000s.

2.4.1 Steel

Steel used in road and bridge construction (as well as construction of related buildings) is alloy steel (not stainless), in different forms, including rods, bars and wire. In table (2.5) I list in red italics the types of steel most used in the construction of transportation infrastructure, by the Harmonized System classification.

Construction steel is made in large part from scrap steel, which is traded in large quantities. Therefore, production is not linked to natural resources availability to such an extent that raw inputs other than labor need enter the model. In addition, there are many different grades of steel, each with their own specialized uses, therefore making CES demand and monopolistic competition a suitable theoretical framework to describe steel manufacturing.

The construction steel categories market in table (2.5) map mainly into Standard International Trade Classification (SITC) revision 3 categories 675, 676, as well as 672 and 678. Figure (2.2) shows how these four categories place in the spectrum of industries in terms of transport costs and differentiation, using the same estimates as in the arms paper: freight rate from Hanson and Xiang (2004) and substitution elasticities from Broda and Weinstein (2006). It appears that, while transport costs are significant, differentiation varies, with the bulk of construction steel (categories 675 and 676) being relatively homogeneous.

As mentioned before, the control industries indicated here and used in regressions are still the ones isolated with respect to military weapons. Figure (2.2) shows that these control goods are on average as homogeneous, or more so, than steel categories, and have significantly lower transport costs.

I estimate equation (2.5), where the construction material is represented by the group of non-stainless alloy steel categories discussed, and public transportation spending stands in for exporters' infrastructure spending I. Results are displayed in table (2.6). The preferred specification dictated by the model is shown in column 3. I also experiment with holding back some controls (columns 1 and 2), or restricting the sample (column 4) in order to separate the effect of sample composition from the effect of adding endowment controls as we switch from the 2nd to the 3rd specification. Column 5 introduces a control for relative GDP, to ensure results are not an artifact of differences in country size. The coefficient on the differenced ratio of transportation spending to GDP is positive and significant across all specifications, indicating strong home market effects. The coefficient in column 3 suggests that a 10 percent increase in transportation infrastructure spending is associated with a 14.9 percent increase in exports of alloy steel (recall that any increase in exports would be evidence of the home market effect).

All other coefficients have the signs we expect: greater distance between countries makes it less likely to ship a heavy product like steel, but being neighbors increases the volume of steel trade; past colonial relationship has a positive effect, which suggests historical bilateral trade relationships in steel; high capital endowment is a plus, but large country area is a (weak) minus. Education and shared language are inconsequential, meaning production and trade in steel require a similar level of skill as the control goods.

2.4.2 Machinery

I have isolated the following three categories of machinery and vehicles most likely to be used in construction projects.

8429	Self-propelled bulldozers, angledozers, graders, levellers, scrapers, mechanical shovels, excavators, shovel loaders, tamping machines, and road rollers.
870510	Crane lorries.
870540	Concrete-mixer lorries.

Construction machinery (HS=8429) maps into SITC category 723. Crane and concrete-mixer lorries (HS=870510 and 870540) map into category 782, which also contains vehicles for the transportation of goods, so it has been omitted from the graph of substitution elasticity σ vs. freight rate. Figure (2.3) shows that construction machinery is a little above average in terms of differentiation, and a little below average in transport costs (4th decile of both σ_{BW} and freight rate).

Regression results indicate that construction machinery (including crane and concrete-mixer lorries) does display home market effects when controlling for capital, labor and education, although the result is insignificant in absence of these controls (see table 2.7).

I interpret these results to mean that, unlike for steel, endowment measures (especially capital and education) are needed here to control for different production capabilities and costs. Machinery and vehicles require a higher degree of know-how than steel production, as well as access to more sophisticated technology. Nonetheless, once these differences are accounted for, domestic demand has a positive effect on exports, in other words the home market effect carries through.

2.4.3 Cement

Cement is widely used in the construction of roads, bridges, and airports. It is captured in the Harmonized System by category 2523: "Portland cement, aluminous cement, slag cement, supersulphate cement and similar hydraulic cements, whether or not coloured or in the form of clinkers." The corresponding 3-digit SITC category is 661: "Lime, cement, and fabricated construction materials (except glass and clay materials)", which includes asphalt. Figure (2.4) shows that cement has median differentiation (5th decile of σ_{BW}), but very high transport costs.

Table (2.8) shows regression results for cement (HS=2523), and table (2.9) restricts the product sample to just Portland cement (HS=252329 - the type of cement most commonly used in road construction). Results are consistent with the idea that cement production is sticky: the relevant coefficient is *negative* and significant: the more a country spends on transport infrastructure, the less it exports cement products.

One possibility to consider is that transport costs are too extreme for the home market effect to hold. Figure 2.5 shows that cement (SITC 661) has the second highest freight rate among all industries for which the data is available. The only category with higher transport costs is *Clay construction materials and refractory construction materials* (SITC 662). I test existence of the home market effect for this industry, by using country size differences as in Hanson and Xiang (2004), and find no empirical evidence that trade patterns are correlated with country size (results not reported). This is consistent with the intuition that goods with prohibitively high transport costs will simply not be traded, and so varying demand will not impact trade in either direction.

However, the relevant coefficients in tables (2.8) and (2.9) are significantly negative, rather than zero, suggesting that high transport costs are not a complete explanation.[4]

To explain a *reverse* home market effect, we need to consider the specifics of cement production in more detail. Cement is produced using non-metallic minerals like limestone, clay and gypsum. These have high transport cost to value ratios, and are therefore not traded in significant quantities. Another major input into cement production is electricity. As these inputs are not traded and have decreasing returns to scale beyond an initial efficient scale, they explain why production does not relocate with changes in demand. Yet another driver of this stickiness in production is the pollution associated with cement manufacturing, which requires somewhat lax environmental regulations. The result is that regions and countries that experience surges in demand will have to respond by decreasing their exports and increasing their imports of cement, consistent with the pattern I am observing.

2.4.4 Asphalt

Asphalt (also known as bitumen) is a good candidate to test the model empirically, because its primary use is in road construction, where it is used as the glue or binder mixed with aggregate (sand and gravel) particles to create asphalt concrete.

Similarly to cement, shipping asphalt is expensive, but not prohibitively so,[5] and its production does display scale economies, which suggests we should expect the home market effect to hold.

However, asphalt is a petroleum product - it is obtained from either natural deposits or as a byproduct of the petroleum industry (petroleum asphalt). Since production is tied to the extraction and processing of oil or other natural deposits, and is further limited by environmental regulations, producers may not be able to relocate in response to demand changes. Regression results in table (2.10) show that, indeed,

[4]Chad Syverson (Syverson, 2004, 2007) discusses how, due to extreme transport costs, the market for ready-mix *concrete* is focused on satisfying local demand. However, *cement* (before mixing into concrete) is not prohibitively expensive to move across large distances, and it is traded internationally.

[5]Recall that asphalt is captured in the wider SITC=661 cement category depicted in figure (2.4).

asphalt exports drop slightly, although not significantly, in response to increases in domestic demand.

2.5 Housing expenditure

Transportation spending far outranks all other types of government infrastructure expenditure, which is what makes it a top pick for this exercise. Another key factor is that different types of government spending are correlated, whether they are substitutes or complements. As an example, housing and transportation infrastructure spending are positively correlated for some countries (Ireland .80, UK .79, Bulgaria .67). For other countries, there is strong negative correlation (Sweden -.82, Poland -.76, Norway -.49).

As I explore the possibility of using another spending category, aside from transportation, the challenge is identifying construction materials that are relevant to it, but which are *not* major inputs into road and bridge construction. For example, cement is a significant input into housing developments, but since it is also widely used in transportation infrastructure projects, and transportation spending is almost one order of magnitude higher than spending housing developments, I cannot claim that housing expenditure is a good demand measure for cement.

One construction material used in housing development but far less in transportation-related projects is glass.[6] I use this fact to test the model using public housing spending as well.

2.5.1 Glass

Table (2.11) lists all products in the Harmonized System *glass and glassware* category. Goods used in construction are marked in red italics.

The SITC category that best corresponds to this group of goods is 664: glass, not including glassware. As figure (2.7) shows, glass is well differentiated and has high

[6]With the caveat that transportation spending includes construction of airports, train stations, etc, which do make use of glass.

(but not extreme) transport costs - making it an ideal candidate for home market effects.

However, results reported in table (2.12) suggest that, like cement, glass clearly displays the reverse of the home market effect. Some of the same arguments discussed for cement apply, as well: high fixed costs mean there is a long lag in capacity adjustment to changes in demand, so in the short to medium run excess demand is accommodated through increased imports. Production involves large quantities of heavy and inexpensive (therefore non-traded) inputs, including fresh water, therefore production locates close to input availability points. Additionally, glass manufacturing processes impact the local environment negatively through noise, water and air pollution[7] - this makes it very costly, time consuming, or even impossible to open factories in some locations, or expand existing ones.

2.6 Conclusion

Construction steel and machinery display home market effects, in agreement with the theoretical model proposed, but cement and glass exhibit the opposite pattern: for these latter industries (as well as perhaps for asphalt, where the result is not statistically significant), higher demand leads to reduced exports. Further consideration of the specifics of production of cement, asphalt and glass suggest that a single-input production model as in the standard monopolistic competition approach may simply not be suitable to describe these industries, which utilize non-traded inputs, are energy intensive and highly polluting. Local comparative advantage appears to play a larger part in establishing production location than geographic variation in demand.

It is also helpful to keep in mind that, if capacity adjustment is slower in certain industries like cement, it may be difficult to tease out the effect of changing demand on long-term production, and standard trade models which assume instantaneous adjustment may again not be suitable.

My results are in agreement with previous estimates by Davis and Weinstein

[7]One source of air pollution is transportation of the dusty inputs.

(1999), who found home market effects for “iron and steel”, “electrical machinery” and “transport equipment”, but not for “stone, clay, and glass”.

In terms of practical short-run implications, the results suggest that we can expect the steel industry to expand as a result of increased domestic infrastructure spending. Conversely, the trade balance will drop when domestic demand for cement or glass spikes. This suggests that, given a choice between stimulus spending in different types of infrastructure projects, the government should favor steel- and machinery-intensive projects if the goal is to increase domestic employment and improve the trade balance.

Acknowledgements

I thank Gordon Hanson, Thomas Baranga, Eli Berman, Lawrence Broz, Roger Gordon for their comments, and Chad Krause for answering my questions on the construction industry.

2.7 Figures

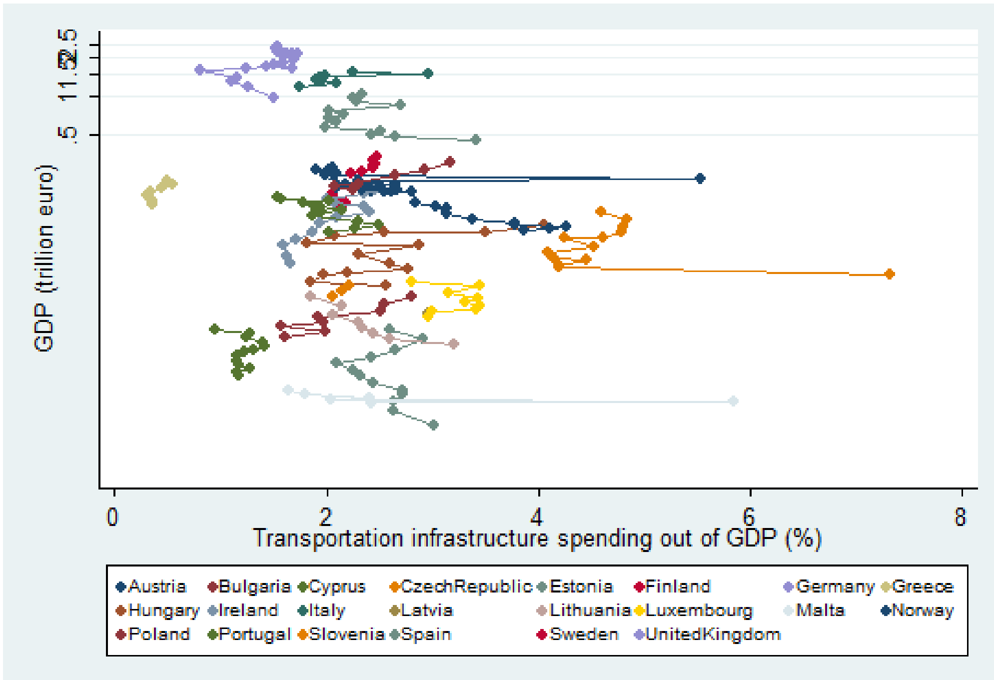

Figure 2.1: Transport infrastructure spending out of GDP vs. GDP

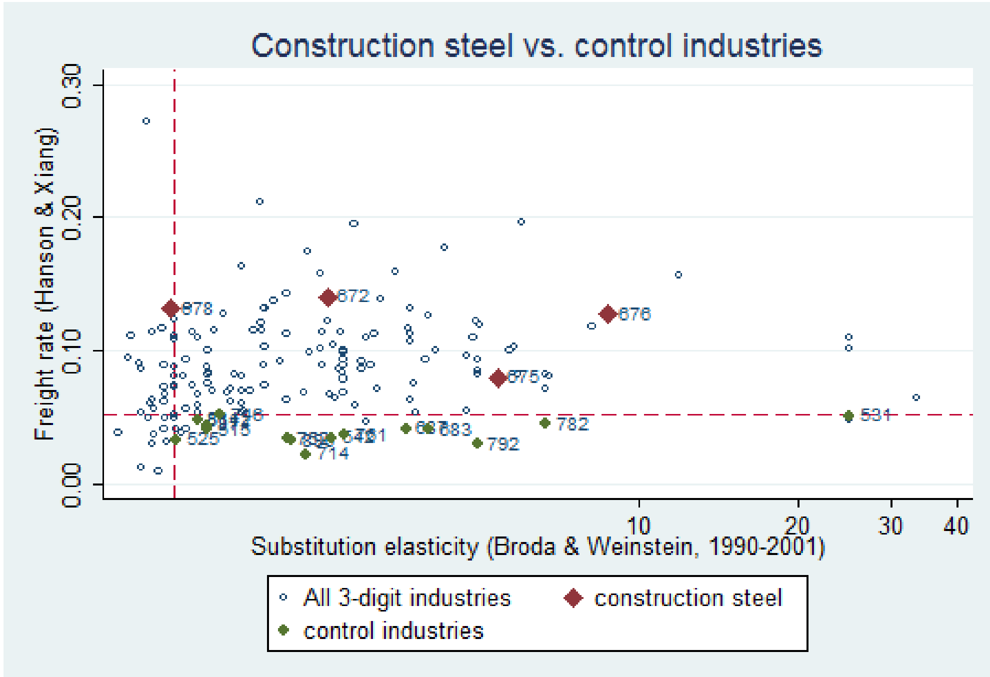

Figure 2.2: Differentiation and freight rate of construction steel vs. other industries

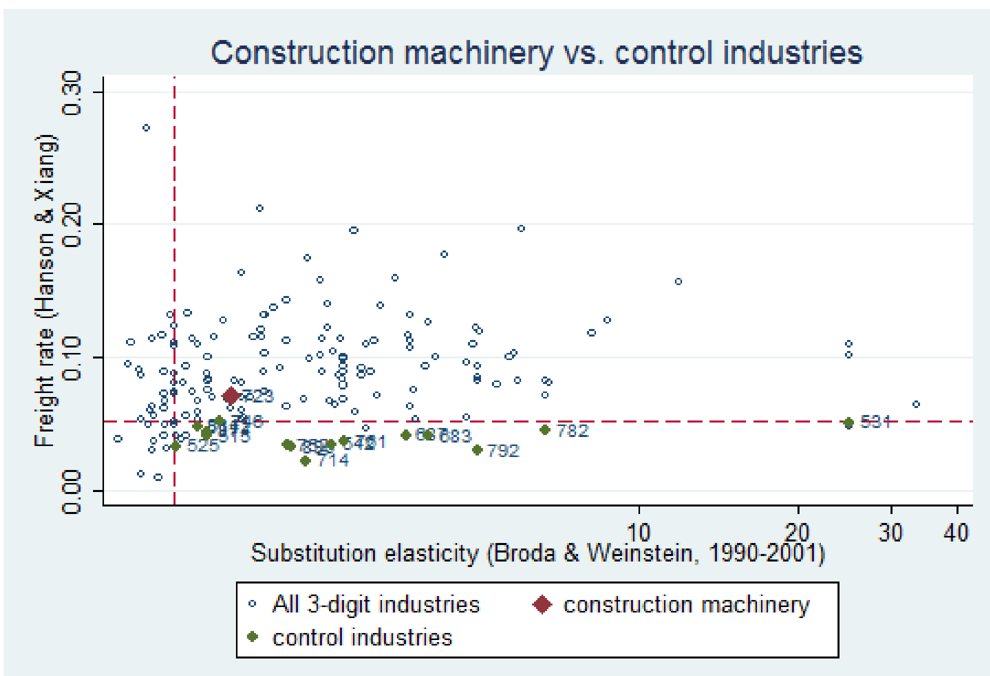

Figure 2.3: Differentiation and freight rate of construction machinery vs. other industries

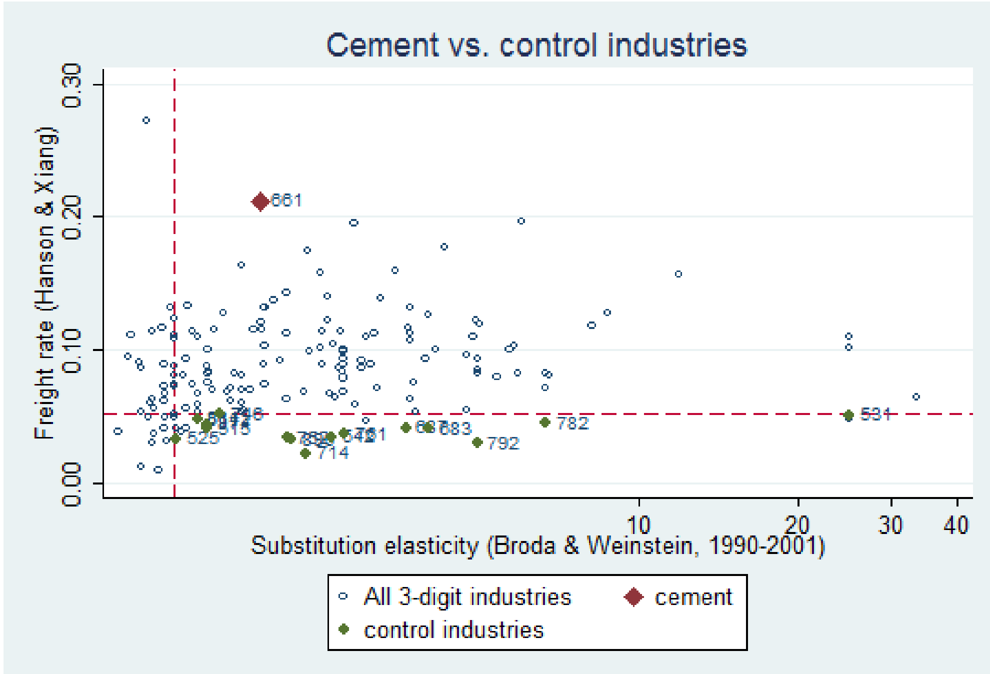

Figure 2.4: Differentiation and freight rate of cement vs. other industries

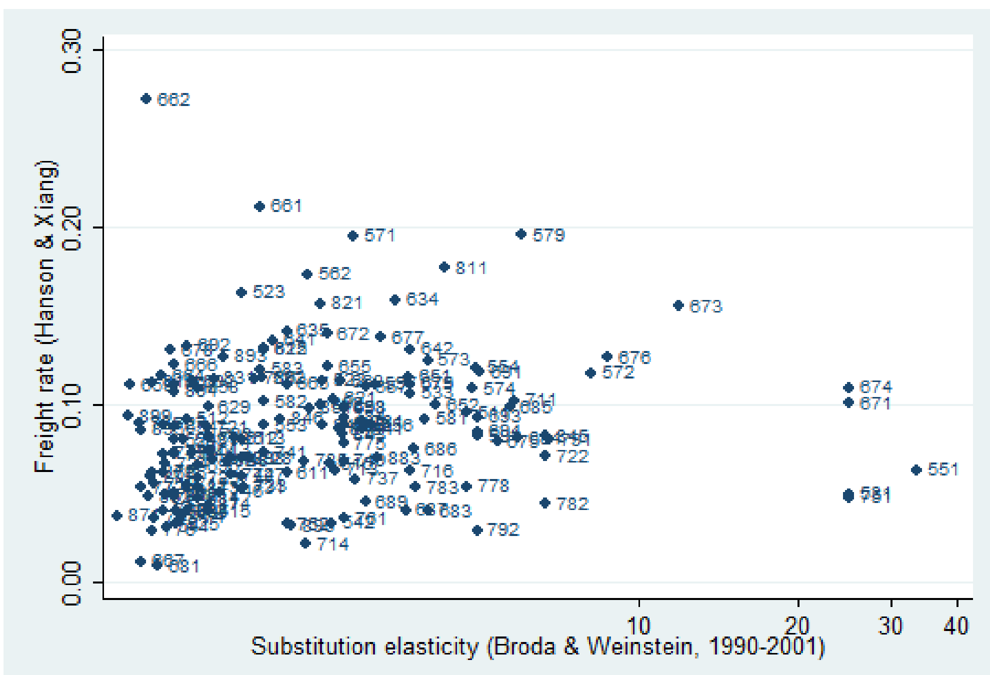

Figure 2.5: Differentiation and freight rate for all industries

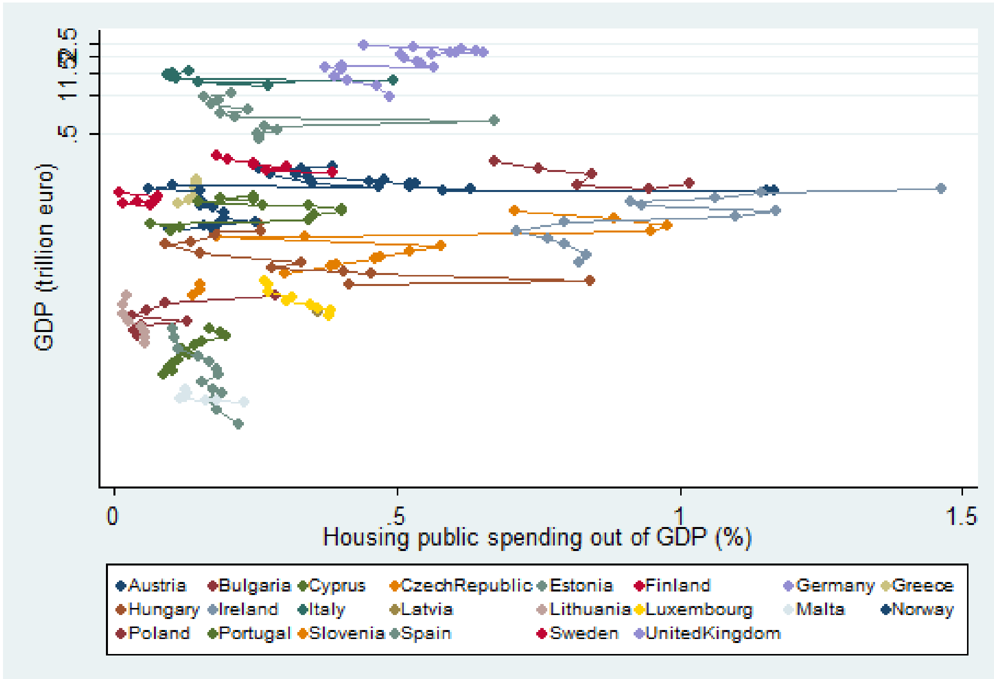

Figure 2.6: Public spending on housing projects out of GDP vs. GDP

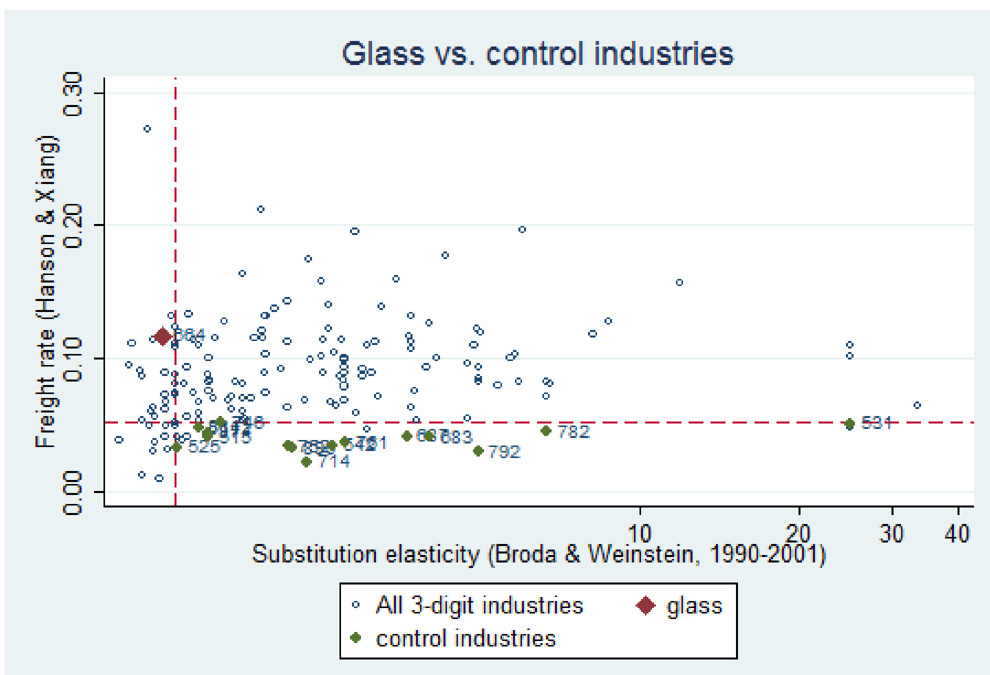

Figure 2.7: Substitution elasticity and freight rate of glass vs. other industries

2.8 Tables

Table 2.1: Government expenditure categories - Eurostat data

01		General public services
02		Defence
03		Public order and safety
04		Economic affairs
	04.1	General economic, commercial and labour affairs
	04.2	Agriculture, forestry, fishing and hunting
	04.3	Fuel and energy
	04.4	Mining, manufacturing and construction Administration, policing of these industries.
	04.5	***Transport*** Includes construction and maintenance of road, water, railway, air, pipeline and other transport systems.
	04.6	*Communication* Includes construction and maintenance of communication systems (postal, telephone, telegraph, wireless, satellite)
	04.7	Other industries
	04.8	R&D Economic affairs
	04.9	Economic affairs n.e.c.
05		Environmental protection
	05.1	Waste management
	05.2	*Waste water management* Includes construction and maintenance of sewer lines
	05.3	Pollution abatement
	05.4	Protection of biodiversity and landscape
	05.5	R&D Environmental protection
	05.6	Environmental protection n.e.c.
06		Housing and community amenities
	06.1	***Housing development***
	06.2	Community development Community planning, excludes implementation.
	06.3	*Water supply* Includes construction and operation of potable water supplies.
	06.4	*Street lighting* Excludes highway lighting
	06.5	R&D Housing and community amenities
	06.6	Housing and community amenities n.e.c.
07		Health
08		Recreation, culture and religion
09		Education
10		Social protection

Table 2.2: Data availability by country

Start year	Countries
1990	Norway
1995	Ireland, Czech Rep, Estonia, Spain, Cyprus, Luxembourg, Hungary, Austria, Portugal
1996	United Kingdom
2000	Bulgaria, Germany, Italy, Lithuania
2001	Greece, Malta, Sweden
2002	Poland, Finland
2005	Slovenia
2007	Latvia

Table 2.3: Expenditure categories and corresponding construction materials

Gov't expenditure category	Transportation infrastructure	Housing development
Candidate materials	Steel Cement Asphalt Machinery	Glass

Table 2.4: Set of control goods

	SITC rev 3 description	freight rate	σ	HS concordance
714	Engines and motors, nonelectric	0.0217	2.37	8411
792	Aircraft and associated equipment; spacecraft	0.0295	4.98	8801-3, 8805
896	Works of art, collectors' pieces and antiques	0.0323	2.23	9701-6
525	Radioactive and associated materials	0.0331	1.35	2844-6
752	Computers	0.0333	2.18	8471
542	Medicaments	0.0338	2.65	3003, 3004
761	TV receivers	0.0364	2.8	8528
683	Nickel	0.0402	4.04	7502, 7504-7
515	Organo-inorganic compounds	0.0404	1.55	2930-2935
687	Tin	0.0409	3.65	8001, 8003-6
874	Measuring and analysing instruments	0.0440	1.55	9014-7,9023-7,9030-3
782	Motor vehicles for the transport of goods	0.0445	6.7	8704-5
514	Nitrogen-function compounds	0.0475	1.48	2921-9
881	Photographic apparatus and equipment, n.e.s.	0.0477	1.48	9006-8, 9010
531	Synthetic organic coloring matter	0.0504	25.03	3204-5
746	Ball or roller bearings	0.0512	1.63	8482

Table 2.5: Iron and steel (HS=72) subcategories

7201-7205	[Primary materials]
7206-7217	[Iron and non-alloy steel]
7218-7223	[Stainless steel]
Other alloy steel	
7224	*Other alloy steel in ingots or other primary forms; semi-finished products of other alloy steel.*
7225	*Flat-rolled products of other alloy steel, of a width of 600 mm or more.*
7226	*Flat-rolled products of other alloy steel, of a width of less than 600 mm.*
7227	*Bars and rods, hot-rolled, in irregularly wound coils, of other alloy steel.*
7228	*Other bars and rods of other alloy steel; angles, shapes and sections, of other alloy steel; hollow drill bars and rods, of alloy or non-alloy steel.*
7229	*Wire of other alloy steel.*

Table 2.6: Alloy steel, other than stainless

	s1	s2	s3	s3-sample	s5
	(1)	(2)	(3)	(4)	(5)
ln(trans/GDP)	1.22	1.08	1.49	1.37	1.31
	(.20)***	(.20)***	(.28)***	(.31)***	(.23)***
distance		-.36	-.45	-.60	-.24
		(.13)***	(.12)***	(.22)***	(.12)**
colonial relationship		-.33	.71	.34	.81
		(.19)*	(.26)***	(.31)	(.25)***
common language		.16	-.31	.02	-.55
		(.25)	(.36)	(.36)	(.33)*
common border		.79	.61	.70	.67
		(.17)***	(.27)**	(.35)**	(.26)***
capital/worker			1.75		2.48
			(.53)***		(.52)***
land/worker			-.54		-1.17
			(.32)*		(.34)***
years schooling			-.008		.13
			(.86)		(.82)
ln(GDP)					-.97
					(.36)***
Obs.	16265	16265	7685	7685	7685
e(N-clust)	226	226	68	68	68
R^2	.06	.08	.2	.15	.22

Notes: dep. variable = $\ln\left(\frac{S_{cjk}/S_{chk}}{S_{ojk}/S_{ohk}}\right)$: flow of construction materials (c) from exporters j and h to importer k, vs. flows of control goods (o). Exporters are ordered so that exporter 1 has higher GDP, and the sample is restricted so that exporter 1 also has higher gov't expenditure. Year and importer country dummies are included in all regressions. Exporter-pair clustered standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***).

Table 2.7: Construction machinery and vehicles

	s1	s2	s3	s3-sample	s5
	(1)	(2)	(3)	(4)	(5)
ln(trans/GDP)	.07	.03	.57	.21	.48
	(.16)	(.16)	(.18)***	(.21)	(.17)***
distance		-.11	-.09	-.35	.03
		(.13)	(.13)	(.22)	(.11)
colonial relationship		.005	.30	-.11	.35
		(.14)	(.22)	(.27)	(.22)
common language		.32	.29	.68	.09
		(.22)	(.29)	(.30)**	(.29)
common border		.44	.41	.44	.45
		(.13)***	(.17)**	(.21)**	(.16)***
capital/worker			1.57		2.07
			(.36)***		(.42)***
land/worker			-.88		-1.32
			(.25)***		(.30)***
years schooling			.60		.73
			(.73)		(.69)
ln(GDP)					-.64
					(.30)**
Obs.	13552	13552	7673	7673	7673
e(N-clust)	227	227	68	68	68
R^2	.03	.04	.15	.07	.17

Notes: dep. variable = $\ln\left(\frac{S_{cjk}/S_{chk}}{S_{ojk}/S_{ohk}}\right)$: flow of construction materials (c) from exporters j and h to importer k, vs. flows of control goods (o). Exporters are ordered so that exporter 1 has higher GDP, and the sample is restricted so that exporter 1 also has higher gov't expenditure. Year and importer country dummies are included in all regressions. Exporter-pair clustered standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***).

Table 2.8: Cement (HS=2523)

	s1	s2	s3	s3-sample	s5
	(1)	(2)	(3)	(4)	(5)
ln(trans/GDP)	-.85	-1.17	-1.40	-1.39	-1.42
	(.29)***	(.26)***	(.31)***	(.34)***	(.30)***
distance		.21	-.09	.25	-.004
		(.15)	(.15)	(.18)	(.18)
colonial relationship		-.02	-.65	-.21	-.62
		(.27)	(.30)**	(.32)	(.29)**
common language		-1.17	-1.32	-1.76	-1.46
		(.33)***	(.32)***	(.35)***	(.30)***
common border		3.02	3.08	3.05	3.10
		(.30)***	(.25)***	(.29)***	(.24)***
capital/worker			-2.05		-1.75
			(.33)***		(.41)***
land/worker			.29		.03
			(.19)		(.23)
years schooling			-1.71		-1.72
			(.61)***		(.62)***
ln(GDP)					-.38
					(.33)
Obs.	5893	5893	3337	3337	3337
e(N-clust)	210	210	67	67	67
R^2	.09	.2	.32	.23	.32

Notes: dep. variable = $\ln\left(\frac{S_{cjk}/S_{chk}}{S_{ojk}/S_{ohk}}\right)$: flow of construction materials (c) from exporters j and h to importer k, vs. flows of control goods (o). Exporters are ordered so that exporter 1 has higher GDP, and the sample is restricted so that exporter 1 also has higher gov't expenditure. Year and importer country dummies are included in all regressions. Exporter-pair clustered standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***).

Table 2.9: Portland cement (HS=252329)

	s1	s2	s3	s3-sample	s5
	(1)	(2)	(3)	(4)	(5)
ln(trans/GDP)	-.17	-.65	-1.73	-1.41	-1.56
	(.43)	(.38)*	(.40)***	(.42)***	(.37)***
distance		.93	.54	.92	.78
		(.22)***	(.26)**	(.27)***	(.29)***
colonial relationship		.34	-.07	-.08	-.05
		(.32)	(.47)	(.49)	(.49)
common language		-.45	-.67	-1.08	-1.06
		(.42)	(.53)	(.56)*	(.54)**
common border		3.70	3.43	3.75	3.36
		(.38)***	(.48)***	(.59)***	(.41)***
capital/worker			-.78		.24
			(.50)		(.63)
land/worker			.80		-.16
			(.36)**		(.39)
years schooling			-2.72		-2.62
			(.69)***		(.67)***
ln(GDP)					-1.15
					(.48)**
Obs.	2461	2461	1353	1353	1353
e(N-clust)	174	174	62	62	62
R^2	.17	.3	.36	.34	.38

Notes: dep. variable = $\ln\left(\frac{S_{cjk}/S_{chk}}{S_{ojk}/S_{ohk}}\right)$: flow of construction materials (c) from exporters j and h to importer k, vs. flows of control goods (o). Exporters are ordered so that exporter 1 has higher GDP, and the sample is restricted so that exporter 1 also has higher gov't expenditure. Year and importer country dummies are included in all regressions. Exporter-pair clustered standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***).

Table 2.10: Asphalt

	s1	s2	s3	s3-sample	s5
	(1)	(2)	(3)	(4)	(5)
ln(trans/GDP)	-.05	-.18	-.36	-.17	-.37
	(.27)	(.26)	(.33)	(.33)	(.33)
distance		.06	-.44	-.36	-.41
		(.19)	(.21)**	(.23)	(.17)**
colonial relationship		-.15	.07	-.01	.07
		(.24)	(.32)	(.31)	(.31)
common language		-.99	-1.07	-.99	-1.09
		(.32)***	(.36)***	(.39)**	(.35)***
common border		1.50	1.60	1.67	1.61
		(.20)***	(.26)***	(.27)***	(.26)***
capital/worker			.52		.61
			(.47)		(.51)
land/worker			.22		.14
			(.25)		(.34)
years schooling			-.91		-.89
			(1.04)		(1.05)
ln(GDP)					-.11
					(.40)
Obs.	7860	7860	4974	4974	4974
e(N-clust)	196	196	68	68	68
R^2	.06	.09	.15	.14	.15

Notes: dep. variable = $\ln\left(\frac{S_{cjk}/S_{chk}}{S_{ojk}/S_{ohk}}\right)$: flow of construction materials (c) from exporters j and h to importer k, vs. flows of control goods (o). Exporters are ordered so that exporter 1 has higher GDP, and the sample is restricted so that exporter 1 also has higher gov't expenditure. Year and importer country dummies are included in all regressions. Exporter-pair clustered standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***).

Table 2.11: Glass and glassware (HS=70) categories

70	Glass and glassware
7001	Glass cullet, waste or scrap, glass in the mass.
7002	*Glass in balls (other than of heading No. 70.18), rods or tubes, unworked.*
7003	*Cast and rolled glass, in sheets and profiles, whether or not having an absorbent, reflecting or non-reflecting layer, but not otherwise worked.*
7004	*Drawn or blown glass, in sheets.*
7005	*Float glass, surface ground, polished glass in sheets.*
7006	*Glass of heading No. 70.03, 70.04 or 70.05, bent, edge-worked, engraved, drilled, enameled or otherwise worked, but not framed or fitted with other materials.*
7007	*Safety glass, consisting of toughened (tempered) or laminated glass.*
7008	*Multiple-walled insulating units of glass.*
7009	Glass mirrors, whether or not framed, including rear-view mirrors.
7010	Glass bottles, flasks, jars, phials, stoppers, etc
7011	Glass envelopes (including bulbs and tubes), open, and glass parts thereof, without fittings, for electric lamps, cathode-ray tubes or the like.
7012	Glass inners for vacuum flasks, other vacuum vessels.
7013	Glassware for table, kitchen, toilet, decoration.
7014	Signalling glassware, unworked optical elements.
7015	Glasses for spectacles, clocks, watches, unworked.
7016	*Paving blocks, slabs, bricks, squares, tiles and other articles of pressed or moulded glass, whether or not wired, of a kind used for building or construction purposes; glass cubes and and other glass smallwares.*
7017	Laboratory, hygienic or pharmaceutical glassware etc.
7018	Glass beads, imitation stones (not jewel), ornaments.
7019	Glass fibres, glass wool, and articles thereof (e.g. yarn, woven fabrics).
7020	Articles of glass, nes.

Table 2.12: Glass used for construction (HS=7002-7008, 7016)

	s1	s2	s3	s3-sample	s5
	(1)	(2)	(3)	(4)	(5)
ln(hous/GDP)	-.49	-.59	-1.04	-.88	-1.02
	(.16)***	(.15)***	(.21)***	(.20)***	(.23)***
distance		-1.05	-1.06	-.65	-1.09
		(.13)***	(.13)***	(.18)***	(.14)***
colonial relationship		-.73	-.43	-.24	-.44
		(.20)***	(.25)*	(.28)	(.25)*
common language		-.02	-.44	-.63	-.41
		(.27)	(.23)*	(.31)**	(.22)*
common border		.85	1.51	1.74	1.50
		(.16)***	(.25)***	(.30)***	(.26)***
capital/worker			-.75		-.86
			(.72)		(.68)
land/worker			-.19		-.09
			(.26)		(.35)
years schooling			-1.14		-1.20
			(1.02)		(1.07)
ln(GDP)					.13
					(.33)
Obs.	23811	23811	9016	9016	9016
e(N-clust)	220	220	61	61	61
R^2	.04	.12	.27	.2	.27

Notes: dep. variable = $\ln\left(\frac{S_{cjk}/S_{chk}}{S_{ojk}/S_{ohk}}\right)$: flow of construction materials (c) from exporters j and h to importer k, vs. flows of control goods (o). Exporters are ordered so that exporter 1 has higher GDP, and the sample is restricted so that exporter 1 also has higher gov't expenditure. Year and importer country dummies are included in all regressions. Exporter-pair clustered standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***).

2.9 Bibliography

Aschauer, D.A., "Is public expenditure productive?," *Journal of monetary economics*, 1989, *23* (2), 177–200.

Barro, R.J. and J.W. Lee, "International data on educational attainment: updates and implications," *Oxford Economic Papers*, 2001, *53* (3), 541–563.

Broda, C. and D.E. Weinstein, "Globalization and the Gains from Variety," *quarterly Journal of Economics*, 2006, *121* (2), 541–585.

Davis, D.R. and D.E. Weinstein, "Economic geography and regional production structure: an empirical investigation," *European economic review*, 1999, *43* (2), 379–407.

_ and _, "Market access, economic geography and comparative advantage: an empirical test," *Journal of International Economics*, 2003, *59* (1), 1–23.

Dixit, A.K. and J.E. Stiglitz, "Monopolistic competition and optimum product diversity," *The American Economic Review*, 1977, *67* (3), 297–308.

Ford, R., P. Poret, Organisation for Economic Co-operation, Development. Dept. of Economics, and Statistics, *Infrastructure and private-sector productivity*, OECD, 1991.

Hanson, G.H. and C. Xiang, "The Home-Market Effect and Bilateral Trade Patterns," *American Economic Review*, 2004, *94* (4), 1108–1129.

Hirakawa, O., "The Home Market Effect and the International Arms Trade," 2011.

Holtz-Eakin, D. and A.E. Schwartz, "Spatial productivity spillovers from public infrastructure: evidence from state highways," *International Tax and Public Finance*, 1995, *2* (3), 459–468.

Morrison, C.J. and A.E. Schwartz, "State infrastructure and productive performance," 1992.

Munnell, A.H., "Why has productivity growth declined? Productivity and public investment," *New England Economic Review*, 1990, *30*, 3–22.

Syverson, C., "Market structure and productivity: A concrete example," 2004.

_ , "Prices, Spatial Competition and Heterogeneous Producers: an Empirical Test," *The Journal of Industrial Economics*, 2007, *55* (2), 197–222.

Chapter 3

The Success of Entrepreneurial Networks: Evidence from Brazil

Abstract

This paper links the network structure amongst initial employees to the performance of a newly founded firm. We use a large employee-employer linked panel data set from Brazil that allows us to track workers across jobs and establish whether new firm employees have prior joint work experience. We use this information to construct a quantifier for network concentration using the Herfindahl Hirschman Index (HHI), and test the impact of network concentration on new firm performance as measured by survival, employment, and wages. We find that new firms with higher network concentrations, i.e. wherein initial employees have worked together previously, are on average larger, have higher wages and survive longer when controlling for industry fixed effects and employees' human capital, demographic characteristics, formal sector experience, and size of parent firms. This association increases with the initial size of the newly founded firm. However, we find a negative relationship between network concentration and initial firm growth. Finally, we look at how the size of an individual's parent firm affects the success of her new entrepreneurial venture and find that small firm experience correlates with better survival rates, but lower employment and average wages at the new firm.

3.1 Introduction

In recent years, policy makers have placed substantial resources into the promotion of entrepreneurial activity, particularly in small business formation. Entrepreneurial activity can create wealth, employment and innovation, thereby spurring economic growth. It is therefore of great interest to both economists and policy-makers to understand what it means for a start-up venture to be successful, and what factors determine this success.

In this paper, we posit that information relating to entrepreneurial success is implicitly contained in network structures. To evaluate this hypothesis, we exploit the network structure relating the prior career experiences of the founders/ employees in a new firm so as to identify characteristics that lead to more successful entrepreneurial ventures. Specifically, we look at how the initial employees in a newly founded firm are linked in terms of previous employment, and how these links affect the new firm's performance.[1]

This exercise is interesting and relevant for a few reasons. The success of any given venture is determined primarily by three factors: the quality of the proposed undertaking, the competence of its execution, and the external environmental factors that govern market conditions (competition and luck). Our paper will be able to shed some insight into the first two of these three factors. We argue that network structures are associated with factors that reflect the quality of an undertaking and how well it is carried out. Both the quality of a venture and its execution are a function of those individuals pursuing it. Ventures are undertaken both individually and in groups. When individuals come together to engage in entrepreneurial activity, they exchange a variety of information on their compatibility, the availability of opportunities and the viability of ideas. Different individuals have differential access to information and resources, and it is the confluence of this information that potential entrepreneurs use to approximate their expected returns (Aldrich, 1999; Shane and Shane, 2004). This paper is concerned with understanding the impact of the choices that individuals make on whom to work initially on the eventual success of that firm.

We outline four main findings, three of which revolve around the importance of initial employee networks: using the the Herfindahl Hirschman Index (HHI) to measure network concentration, we find a strong link between the HHI and various measures of success. Firstly, we find that survival probabilities increase with network concentration for the average newly founded firm in Brazil. This is to say, new firms wherein the initial employees had worked together previously are more likely to survive. This effect of the HHI on survival is more pronounced for firms that are larger at inception. Secondly,

[1]Ideally we would like to be able to identify the founders of a firm and evaluate how these founders were linked in terms of previous employment. We are however not able to get clean identification to distinguish between founders and employees in our data, The Brazilian RAIS data.

we look at the relationship between the HHI and the new firm's average earnings and employment levels. These measures are intended to proxy for firm revenue which we cannot isolate in our data. We find that firms with higher HHI have both higher average wages and more employees in all periods for which we can track performance (up to six years after entry). Thirdly, when looking at the growth rates of average earnings and employment, we find these measures to be negatively related to the HHI initially and find no effect after three years. Firms wherein new employees are more closely linked via previous employment (i.e. more concentrated networks) are more likely to survive, are larger, and have higher wages. However, they have lower wage growth between inception and the third year. This relationship between network concentration and growth dissipates between years three and six.

The reader should note that we are careful not to make any causal claims as such effects are difficult to isolate in the empirics. Therefore, we propose an assortment of interpretations of our results in section 3.4.3 and suggest avenues for further research that may get at identification in section 3.4.4.

The results described above contribute to various literatures. While most of the empirical literature focuses on the impact of networks on entrepreneurial entry (Nanda and Sorensen, 2009; Giannetti and Simonov, 2009), this is one of few papers to look at the impact of networks on entrepreneurial outcomes (see also Elfring and Hulsink, 2003; Lerner and Malmendier, 2011). We complement this literature by showing that networks affect not only participation in entrepreneurship, but also the success of new ventures.

This paper also supports the theoretical literature on network effects on outcomes. Drawing mainly on the ideas developed in (Granovetter, 1983), we use the HHI to quantify the strength of ties, and assess their impact on entrepreneurial outcomes. In this seminal piece of work, Granovetter notes that there are three possible types of ties linking individuals: strong, weak and absent. Strong ties link individuals within the same social circle and weak ties connect individuals across social circles. While strong ties motivate individuals to be of assistance and work well together, weak ties provide individuals with access to information and resources beyond those available in their own social circle. This concept is taken to an entrepreneurial setting by Rauch and Watson (2007). They argue that when a potential entrepreneur's default option is to interact with readily available individuals within their social network, those who select into involvement with unrelated individuals are willing to incur a higher search cost. This can be interpreted as a signal of higher venture quality. On the other hand,[2] a counterforce to the mechanism proposed by Rauch and Watson is that previously linked individuals are more likely to succeed as the riskiness of the venture is reduced since the founders would all be aware of each other's capabilities and have better information on their joint compatibility, partner calibre and therefore success probabilities. In light of these theories, our empirical findings can be interpreted[3] in the following manner: the finding that network concentration enters positively and significantly into determining survival implies that peer assessment of ability and compatibility helps to ensure that a firm will survive through the initial phases. Starting a firm with closely linked individuals reduces many basic uncertainties, thus allowing the firm to get past the most fragile initial stages. However, conditional on surviving, having access to different resource pools and a higher match quality a la Rauch and Watson (2007) could increase the entrepreneurial potential of a firm.[4] This is reflected by the negative and significant relationship between network concentration, and employment and wage growth in the initial periods. This interpretation is consistent with the notion that the strength of ties does matter, but it enters differently in determining different measures of success. In section 3.4.3 we provide a few other interpretations that are consistent with our results. These interpretations will address three specific alternative channels for our findings. They include selection into different network structures based on underlying qualities,

[2]This is a point we propose that addresses the assumption in Rauch and Watson (2007) that interacting within one's network is the default option

[3]One of a few interpretations, as described in section 3.4.3.

[4]One caveat here is that we measure concentration using all employees, as opposed to just founders since our data does not distinguish between the two. As firms get larger, we capture less of the network effects amongst entrepreneurs since the founders and employees are now less likely to be the same individuals. Further work to disentangle the impact for founders specifically is discussed in the last section of this paper.

access to financing and the relationship between network structure, and the choice to switch industries.

A key feature of our results is that different measures of entrepreneurial success result in different outcomes.[5] Furthermore, our results indicate that fewer covariates enter significantly in determining performance beyond the third year. In thinking through the determinants of entrepreneurial success, it is clear that much consideration should be given to how success is defined. Do we think of successful firms as ones that survive, or ones that grow? And for how long are the standard determinants of success relevant for entrepreneurs?

The fourth main finding of this paper relates entrepreneurial success to parent firm size. We find that small firm experience of the founding employees is associated with higher survival likelihoods for the new firm. However, large firm experience is predictive of higher initial employment and average wages, as well as higher employment growth for the first three years. These findings contribute to the literature on the genesis of entrepreneurs (see Gompers, Lerner and Scharfstein, 2005; Hvide, 2005; Lazear, 2004). This literature seeks to understand where entrepreneurs come from and how this affects their performance. Theoretically, it is conceivable that parent-firms of different sizes produce different types of entrepreneurs for several reasons. Small parent firms, with flatter hierarchical structures, may expose their employees to various layers of the firms' organizational structure in addition to honing their specific roles. This varied skill set creates employees that are "jacks of all trades" (as in Lazear, 2004) who may make more successful entrepreneurs. On the other hand, larger firms may have the resources to fund research intensive projects, thus attracting highly educated and innovative individuals (Gompers, Lerner and Scharfstein, 2005).[6] Given that these large parent firms may not have the ability or desire to internalize all the ideas generated

[5]While looking through the results described in section 3.4 the reader should be careful to note how various standard determinants enter differently depending on the measure of success.

[6]Hvide (2005) reasons, under the assumption that workers select blindly into employment at large or small firms, that larger firms will produce higher quality entrepreneurs, since in these firms wage setting is less fine tuned.

by their employees, this could induce workers to leave and become entrepreneurs so as to realize their ideas. These high ability individuals may then prove to be superior entrepreneurs. Our results lend support to both theories. Employees from small parent firms are better able to ensure the survival of their new firms. Presumably, this involves being able to efficiently manage various different components of the organization, which would be a skill small parent firms expose individuals to. However, large parent firms seem to spawn new firms that have an initial size advantage and ultimately outperform other surviving new firms. This result goes hand in hand with the theory proposed by Gompers, Lerner and Scharfstein (2005) that innovative and highly skilled individuals come from large firms which cannot internalize their ideas, or compensate them to appropriately reflect their true marginal productivity, as suggested by Hvide (2005).

None of the topics identified above are intended to imply anything causal. It is important to note that we will not be able to observe, and hence discuss, whether it is selection into the networks and firms, or the inherent nature of these institutions and organizations that drive outcomes.[7] This paper provides a documentation of the patterns that are observed empirically, so as to fertilize a ground for further research on the mechanisms that could lead to these observations, and potential identification strategies.

The rest of the paper is organized as follows. Section 4.3 describes the data and outlines the estimations strategy we propose. Section 3.3 describes the informative case studies of survival rates for two- and three-employee firms. Section 3.4 presents and discusses the results as well as analyzes the problems plaguing the current specification and section 3.5 concludes.

[7]Endogeneity issues are discussed in section 3.4.4.

3.2 Data and estimation

3.2.1 Dataset

We employ the Brazilian dataset RAIS (*Relação Anual de Informações Sociais* of the Brazilian labor ministry *MTE*), which is an annual census of salaried employees, required to be filled by all employers.[8] Observations are at the job-year level, and contain unique identifying codes for the firm, the establishment (an establishment is a sub-unit of the firm, such as a plant or office), and the worker. This allows us to track employees as they move from one job to the next, and hence to identify employee networks and how they relate to the performance of new firms.

A job observation in RAIS is identified by the employee ID, the employer's tax ID (CNPJ), and dates of job accession and separation. To avoid double-counting employees at new firms, we keep only one observation for each employer-employee pair, choosing the job with the earliest hiring date. If the employee has two jobs at the firm starting in the same month, we keep the highest paying one. The rules on tax ID assignments make it possible to identify new firms (the first eight digits of the tax ID) and new plants within firms (the last six digits of the tax ID). Our data include 71.1 million employees (with 556.3 million job spells) at 5.52 million plants in 3.75 million firms over the sixteen-year period 1986-2001 in any sector of the economy. We limit our attention to the years 1995-2000 to ensure that firms we label as new have not operated before and to be able to track new firms for at least 1 year.

3.2.2 Estimation strategy

A firm's success is measured using survival, employment and wages. [9] We seek to understand how the founders' common history (network relationship) affects the

[8]Actual coverage is estimated to be well above 90% of all formally employed individuals.

[9]Unfortunately, RAIS does not contain any data on revenues or profits. Therefore, we use the combination of employment and wages to proxy for firm performance in terms of revenue.

performance of a firm. In the RAIS dataset, we are unable to reliably isolate the founding members of a firm, therefore we use a measure of the relationship between all the initial employees. Since most of the new firms in the dataset are relatively small,[10] this measure should pick up to a high degree the relationship among founders. Even though we will not be able to assign causal attributes of network links amongst founders, we are still able to identify how the patterns linking employees correlates with firm survival. One could make the argument that in small firms, all, or at least a large fraction of employees play a significant role in determining the success of the firm. As such, understanding how they all relate to each other may be just as relevant as looking at the relationship between owner-entrepreneurs. An alternative strategy would be to simply restrict the sample to only very small firms.[11]

To proxy for network concentration, we use the Herfindahl-Hirschman Index (HHI) to measure the degree of clustering within a firm. A higher HHI value is indicative of a higher degree of commonality among initial employees. The index is computed as follows: $\sum_{j=1}^{J} {s_j}^2$, where s_j is the share of initial employees in the new firm from parent-firm j. Naturally, we exclude all one-employee firms from our analysis.

There are three versions of this index currently in use, depending on the treatment of workers who cannot be traced to a previous job. These untraceable workers could be employees that are new to the workforce, or individuals who have operated exclusively in the informal sector (estimated at 40% of the Brazilian economy) during the previous years of data.

Suppose a firm has N employees, of whom N_p can be traced to a previous job, and N_u cannot. Further assume that these employees have a total of J parent firms, each represented by N_j employees, so that $N = N_p + N_u = \sum_{j=1}^{J} N_j + N_u$.

Then the first version of the index, HHI_1, counts all employees in the

[10] 91.6% of new firms have 10 or fewer employees while 96.4% of new firms have 20 or fewer employees.

[11] One caveat here is that there may just be 1 founder/entrepreneur even in these very small firms. Hence, it is not clear that we will be able to make claims about the relationship amongst founders even in these cases.

denominator, but assigns zero shares to those that cannot be tracked. The second version is computed using only trackable workers' information (and only for firms with at least 2 such employees), while the third assumes that all workers that cannot be tracked come from different parents:

$$\begin{aligned} HHI_1 &= \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2 \quad \text{for } N \geq 2 \\ HHI_2 &= \sum_{j=1}^{J}\left(\frac{N_j}{N_p}\right)^2 \quad \text{for } N_p \geq 2 \\ HHI_3 &= \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2 + N_u\left(\frac{1}{N}\right)^2 \quad \text{for } N \geq 2 \end{aligned}$$

Version 2 is the cleanest in terms of assumptions made, but since it requires that we have at least two trackable employees, it reduces the sample size significantly. Version 3 conforms quite well to the data (as we'll show using 2-and 3-employee firm statistics in the next section) and is our preferred measure.

One drawback of version 3 (as well as version 2) is that the range of values it can take depends heavily on the number of initial employees: in a 2-person firm, HHI_3 has a minimum value of 0.5 when the two employees have no shared experience. However, as the number of employees N gets large, HHI_3 can take on values very close to zero. As such, we create a new rescaled index, HHI_4, to facilitate interpretation of results. In re-scaling this index, regardless of firm size, a value of 0 means there is no joint network experience while a value of 1 means that all employees were previously linked via employment.

HHI_4 is defined as follows: starting from our preferred index measure HHI_3, we subtract the minimum value ($1/N$) it can take for a given number of employees, then rescale the new measure so that the maximum value is again 1. $HHI_4 = \frac{(HHI_3 - 1/N)}{(1-1/N)}$, which simplifies to:

$$HHI_4 \quad = \quad \sum_{j=1}^{J} \frac{N_j(N_j-1)}{N(N-1)} \qquad \text{for } N \geq 2$$

Where once again, j indexes parent firms. Note that, since in a group of n individuals there are $n(n-1)/2$ possible pairs, HHI_4 can be interpreted as the share of all pairs of employees that are joined by common work experience. Table 3.4 shows average values of HHI_3 and HHI_4 for different initial firm sizes. HHI_3 decreases monotonically with the number of initial employees, with the exception of $N > 100$.[12] Comparing HHI_3 and HHI_4, we note that, as intended, HHI_4 is less volatile across initial firm sizes, therefore serving as a better measure for comparisons across initial size categories. In the appendix, we report survival regression results for HHI_1, HHI_2, and HHI_3 as well for completeness.

We regress firm survival on the HHI index and other relevant variables for each cohort from 1995 through 2000, in a linear probability model.

$$y_{it} = \beta_0 + \beta_1 HHI_{i,0} + \beta_2 X_{i,0} + \varepsilon_{it} \tag{3.1}$$

y_{it} is the measure of success of firm i at time t.[13] This can take one of three forms: a survival indicator, performance in terms of employment, and performance in terms of average wages. For the latter two measures, we examine performance both at entry, and up to 6 years later, conditional on survival. Additional controls $X_{i,0}$ include: share of initial employees that can be tracked to a previous job, cohort and sector fixed effects, initial firm size controls, and employee experience and qualifications (average employee age, education level and wage at parent firm).

[12]This suggests that as the firm size becomes large enough, we may be capturing some divestitures.
[13]Alternatively, we can limit the sample to firms that survived at $t-1$.

3.3 Survival case studies

We include a brief analysis of the average survival rates by common history of 2 and 3 person firms started between 1995 and 2000. These entrants make for an interesting case study, since they can have only a handful of possible HHI indices.

Survival of 2-employee firms

Figure 3.1 shows survival rates for firms with just 2 initial employees using HHI_1. This index takes four possible values, while HHI_2, HHI_3, and HHI_4 can each have two values, as shown in table 3.1 (however, HHI_2 is undefined whenever fewer than two employees can be tracked to previous employment).

Figure 3.1 indicates that using the HHI_1 measure of concentration, survival is very similar across index values 0, 0.25 and 0.5. This suggests that two-man firms wherein one or both employees cannot be tracked behave similarly (in terms of survival) to two-man firms wherein both employees can be tracked, but to different parents. This observation is consistent with the assumption behind versions 3 and 4 of the HHI index: that whenever an employee cannot be tracked to a previous formal sector job, we can reasonably assume they do not have joint work experience with any other employees at the new firm.

Figure 3.1 also shows that firms with a HHI of 1 display a significantly higher survival rate. This graph confirms that 2-employee firms wherein both employees had common prior work experience are more likely to survive then firms whose two employees were not observed to work together before.

Survival of 3-employee firms

The same qualitative observations hold when we repeat the exercise for firms with 3 initial employees. Table 3.2 shows the possible index values in the 9 possible scenarios.

Figure 3.2 shows that firms in the first category with HHI_1 values of 0, 0.11, 0.22, and 0.33 all perform similarly and are less likely to survive than firms with at least 2 workers who share a common parent. Firms with 2 workers with common history have survival rates that are higher and are similar regardless as to whether the third employee is trackable. Finally, firms where all 3 employees share a common parent do significantly better than those in the other two categories. This again indicates that workers who cannot be tracked enter in an almost identical fashion to workers from different parent firms. This confirms the rationale for our use of HHI_3 and its rescaled counterpart HHI_4, over HHI_1 and HHI_2, since these indices give us exactly 3 possible values for the HHI, corresponding to the 3 groups of firms that perform similarly.

3.4 Results

We describe results for different measures of new firm success: first, using survival (section 3.4.1), and then employment and wages (3.4.2). Section 3.4.3 provides a detailed discussion of the various ways to interpret the combined results.

3.4.1 Regression results: survival

We run linear probability model regressions for all four versions of the HHI. Our baseline results are computed for HHI_4 and are reported in table 3.6; the results for HHI_1, HHI_2, and HHI_3 can be found in the appendix. For each version of the HHI, we control for 4-digit sector, cohort, initial size category[14], share of the new firm's employees that can be tracked, mean employee age, years of schooling, previous wage, and average parent size category (including a separate category for new firms with no trackable parents).[15]

Common work experience

[14] Categories are: 2 initial employees, 3-4, 5-10, 11-20, 21-50, 51-100, and >100 employees

[15] Although previous wage is included, thus limiting the sample size, some entrants still have unknown parent size, because parent size is measured during the year immediately before the new firm's entry.

Table 3.6 shows that, even controlling for industry fixed effects and employees' human capital, network concentration (as measured by a higher HHI index value) has a positive and significant impact on firm survival. This result holds robustly across all three versions of the HHI index (see the Appendix).

Using the HHI_4 result at $t+6$, we find that a firm whose workers all have previous common experience ($HHI_4 = 1$) is about 8.8 percent more likely to survive than a firm wherein no two workers can be tracked to the same employer ($HHI_4 = 0$). This result provides evidence for the hypothesis that individuals who are better acquainted prior to engaging in entrepreneurial activity have a better capacity to judge the potential of a proposed venture and perhaps work better together, thus resulting in a higher likelihood of success. We are, however, not able to parse out whether it is the mechanism described above, or if this is a function of easier access to financing amongst friends, for instance, that drives this result.

Size of parent firm

We also find that prior experience at small firms is correlated with a higher chance of survival for a new entrant. This result is consistent with Lazear's (2004) notion that entrepreneurs should be "jacks of all trades". Having experience at a smaller firm more likely exposes employees to the different aspects that underly the mechanics of a firm. This mechanism suggests that a small firm veteran will be a more capable new-firm employee since, aside from being qualified in her own specific tasks, she may have exposure to the logistics involved in running a business. Whether this is a pure treatment effect, meaning only the experience itself distinguishes former small firm vs. large firm employees, or whether there is selection into small firm employment by workers who either consciously seek to gain small firm experience, or who simply are a better fit for small scale establishments, is beyond the scope of this paper to disentangle.

In table 3.3 we provide a breakdown of average parent firm sizes in our sample, which consists of new firms of two or more initial employees, started between 1995 and 2000.

Employee human capital

We control for employee human capital by introducing average age, education, and wage at the previous place of employment. As would be expected, we find that a higher average level of education among employees improves the survival odds for a new firm. Controlling for previous employment and education, we find that employee age has a negative impact on firm survival. This finding is interesting since intuitively one could hypothesize how this effect could go either way. One explanation is that perhaps younger employees, despite having less experience, are more creative and adapt better to technology. While empirical evidence in the US shows that older individuals are more likely to become entrepreneurs, our results show that conditional on becoming one, younger individuals perform better. There are many other explanations for this observation, including that some firms with older employees may exit due to founders' retirement rather than poor performance, and our finding should incite further thought on the issue.[16]

Previous earnings are arguably the most direct measure of an employee's human capital. Not surprisingly, average previous wages of initial employees are predictive of better survival odds for the new firm. The effect is highly significant for up to five years after entry. The drop in significance at t+6 is most likely due to the change in cohort composition between column 6 of table 3.6 and the previous regressions.

Transition from informality

The results indicate that having a higher share of initial employees with no known work history is negatively correlated with firm performance (see coefficients on *share trackable* and *no known parent*). If anything, this coefficient is biased towards positive values by the assumption built into HHI_4 that workers coming in

[16]Previous empirical work has found that, even in the US, the average self-employed individual is older than the average wage-employed individual. This could simply be due to older individuals having better access to financing (including from their personal wealth), so they have an easier time starting a new firm, whereas young would-be entrepreneurs only get financial backing if their idea is exceptionally good.

from informality don't share a work history with each other. These individuals could either be new entrants into the labor force or individuals moving from the informal sector into the formal sector. If we find that it is mainly comprised of the latter,[17] we will then be able to comment on the nature of the quality of experience gained in the informal sector versus the formal sectors. Given that this is a particularly salient issue in less developed countries, this would be a useful bite of information to provide.[18]

Initial firm size

Table (3.6) includes controls for initial firm size - specifically for the seven bins indicated in table 3.5. The coefficients, not reported due to space constraints, indicate that larger new firms have significantly and monotonically higher odds of survival. This finding is consistent with common sense expectations that firms which are larger at inception are typically more ambitious ventures, which are better financed. Furthermore, being able to get more individuals immediately on board, either to finance or simply join the venture, is a positive signal for the promise of a firm.

The more interesting question is whether network concentration is correlated with firm survival differentially by starting size. We replace the initial size bins with a continuous measure: the natural logarithm of initial firm size, and add an interaction term between log size and HHI. Results, as reported in table (3.7), show that the interaction term is positive, in other words network concentration has a stronger positive effect for larger firms. We combine the reported coefficients on HHI_4 and the interaction term to find the overall impact of network concentration on the survival odds for firms of different initial sizes. For example, at t+3, a 2-person firm is $4.70+3.06*\ln(2)=6.8$ percent more likely to survive if its founders have worked together before. A 10-person firm is $4.70+3.06*\ln(10)=11.7$ percent more likely to survive at t+3 if all 10 initial

[17]And we could do this by looking at the age of these untracked individuals, especially given the very high labor force participation rate of prime aged males in Brazil. If the entering individual is older, we could assume that on average they are transitioning in from informality into formality rather than new entrants into the labor force.

[18]Upcoming versions of this paper will include these findings.

employees have worked together before, than if none of them had joint prior work experience. This suggests that, as the number of founders gets larger, being from the same group and working well together matters even more than for small initial firms; or perhaps weak ties (i.e. access to a variety of information, markets, financing etc.) become less important as there is a larger starting base.

We also run the specification from table 3.6, excluding initial size controls, for new firms of different sizes. We plot the relevant coefficient on HHI_4 versus entrant size in figure 3.3. This shows how larger firms benefit more from strong ties between the founders, up to a point: for firms of over 50 initial employees, network ties matter less. However, in this size category, true entrepreneurial ventures may no longer compose the majority of new firms, as a significant portion of entrants are likely employer-initiated divestitures and spinoffs, which are not the focus of this paper.

3.4.2 Regression results: firm performance

The results discussed thus far involve survival as the sole measure of success. We now consider other measures of performance, namely employment and wages, in both initial logged values and growth rates. Table 3.8 reports regression results for firms' performance during their first year, as well as 3 and 6 years later. There are two dependent variables: natural log of number of employees and natural log of average wages. We chose employment because job creation is one of the main stated benefits of entrepreneurship, and average wages because we wanted a measure of the amount of economic activity the new firm creates, but wanted it to be disaggregated from the employment effect.

Table 3.9 isolates initial levels from subsequent growth: columns 3 through 6 have been replaced with growth in employment and wages, first from t to $t+3$, and then from $t+3$ to $t+6$. To make the comparison more salient, we limit all regressions reported in this table to a common sample of firms: the 1995 entrants that survived through the last year in our sample (2001), and which have at least some employees at the end of years t, $t+3$, and $t+6$.[19] In addition, growth regressions in table 3.9 (columns 3 through 6) include controls for initial firm size.[20]

Common work experience

Firms with higher network concentration have a higher number of employees (larger firm size) and higher average wages. This relationship is positive and significant in all periods considered: at t (inception), $t+3$ and $t+6$. However, the magnitude of the effect of HHI on both performance measures decreases in time (see table 3.8). To see this initial and growth effects decomposed, consider table 3.9: columns 3 and 4 indicate a negative relationship between HHI and firm performance (conditional on survival) in the first 3 years. Columns 5 and 6 suggest that the effect dissipates beyond the 3 year mark.

The magnitude of these effects is economically significant: table 3.9 results for instance, apply to firms that entered the labor market in 1995 and had employees through 2001. According to columns 1 and 2, new firms with strong employee ties have an immediate advantage in terms of size and wages: holding everything else constant, a new firm A whose employees have all worked together before has 55 percent more employees and 9 percent higher average wages than a new firm B whose employees have no prior joint work experience. In columns 3 and 4, compare firm A above to a firm C of same initial size, same industry and employee characteristics, but no existing employee networks. Firm A will then have 12 percent slower growth in employment, and 3 percent slower growth in average wages during its first 3 years.[21] Firms A and C will have comparable growth in the subsequent 3 years.

[19]Employment and wages in all performance regressions are considered for end-of-year employees only, to avoid double-counting. See Hirakawa, Muendler and Rauch (December 2010) for more details on the way employment spells and earnings are reported in *RAIS*.

[20]We added these to ensure that the effect estimated isn't due merely to compositional effects: for instance, column 1 indicates that more concentrated firms start up larger than new firms with weaker employee networks. But larger firms have lower growth rates. So do more concentrated new firms have lower growth rates, even when compared to other new firms of similar starting size? Column 3 in table 3.9 suggests that the answer is yes.

[21]We also ran the growth regressions from table 3.9 without initial size controls, in effect comparing firms A and B mentioned above. These results, unreported in the paper, suggest firm A will grow at a

Size of parent firm

As we examine initial performance and new firm growth, we find that small firm experience is no longer an advantage, in contrast with results from the survival analysis. Medium and large firm experience is predictive of higher initial employment and average wages, as well as higher employment growth for the first three years (see table 3.9). There is no significant correlation between parent size and wage growth or employment growth beyond the first three years, but early differences persist through the entrants' first six years, conditional on survival, as columns 5 and 6 in table 3.8 show.

Employee human capital

Referring to tables 3.8 and 3.9, the impact of education on employment and average wage is mixed. Education has no detectable relationship to initial firm size and enters positively into employment growth in the first three years. While education is positively related to wage levels, it enters negatively in the estimations wherein average wage growth is the measure of success. Conditional on surviving up to three years, education has no impact on the growth rates of both average wage and employment. To summarize, conditional on survival, the impact of education on entrepreneurial success as measured by growth rates in average wage and employment is unclear.

High average human capital of employees, as indicated by previous wages, is positively associated with initial firm size and average wages. This positive relationship persists as we examine growth performance in terms of employment, but is reversed for growth in wages: firms with higher average previous wages experience slower wage growth (although the overall effect is still of higher wages up to 6 years after entry, once we factor in initial levels). One possible explanation is that previous wages proxy not just for the employees' level of human capital, but also for their opportunity cost. Hiring someone away from a well paying job means you have to offer a higher initial

---

20 percent slower pace than firm B in terms of employment, and at a 4 percent slower rate in terms of average wages.

wage, at the expense of slower wage growth.

Transition from informality

The results using employment and average wage levels also (as with the survival regressions) suggest that having a higher share of initial employees with no known work history is negatively correlated with firm performance. However, we do not find any significant relationship between traceable work history and, employment and wage growth. Conditional on survival, having no known work history does not seem to impact firm growth.

3.4.3 Interpreting the results

Our main results are that network concentration is positively correlated with new firm survival, initial firm size and initial average wage levels. However, network concentration is negatively correlated with new firm growth between inception and the third year, both in terms of employment and average wage. This relationship goes to zero between years three and six. Below we provide four possible mechanisms that would generate these results. Distinguishing between these mechanisms is beyond the scope of this paper, but we hope to provide the reader with some food for thought.

Networks affect entrepreneurial outcomes

Strong network ties are indicative of better information on partner compatibility. Individuals who are well acquainted are also more motivated to ensure their joint survival. Therefore, new firms displaying stronger network links amongst employees are more likely to survive, especially through the initial phases where compatibility and determination to pull through matter most. Moreover, a closer relationship amongst new firm employees may lead to higher initial investments resulting in higher initial wages and a larger initial firm size. This is consistent with HHI being positively related to new firm survival, initial firm size, and initial average wage.

Conditional on survival, access to different (and complementary) information sets and skill backgrounds makes for more entrepreneurial firms that display higher growth in employment and earnings. This is reflected in the negative relationship between *HHI* and growth rates in the first 3 periods.

Financing

When starting a firm with individuals that are well acquainted, the initial level of trust may be higher. Therefore individuals may be more willing to pool resources to be better financed. Alternatively, individuals who are well acquainted may be willing to provide joint collateral, and the shared liability may make banks more willing to finance ventures involving such individuals. These better financed new firms may start larger with higher initial wages and may be more likely to survive, especially given that access to capital is one of the biggest barriers faced by new firms.

However, conditional on survival, once a venture displays good prospects, individuals who are involved may themselves be willing to put in more of their personal/ family wealth. Furthermore, banks may also be more willing to lend once some information on venture quality is revealed. As such, these firms may display higher growth with time. Individuals who are weakly linked may be more likely to have experiences consistent with this explanation. The coefficients on the *HHI* could be a reflection of this differential access to financing which passes through network concentration.

Selection into entrepreneurial networks

Individuals of different types may select into network types differentially. For example, less risk averse individuals (say younger people) may be more likely to work with people they are less acquainted with in exchange for a higher payoff upon success. This characteristics may also reflect the choice to engage in riskier ventures that are more likely to fail. Conditional on initial survival, these firms could be the ones to experience higher growth. This story would also result in the relationships we find between *HHI* and our various measures of success.

Selection by entrepreneur type need not be dictated by risk tolerance alone. Work style preferences and life goals may vary as well, and may account for selection not just into different types of networks, but also into different types of (previous) employment: perhaps there exist two large classes of entrepreneurs: on one hand, there are small scale entrepreneurs who start a firm because they value their independence, not because they aim to create a multi-billion dollar company. They are likely to partner up with close friends, and have most likely been employed at other small firms before. They keep their new business afloat even if performance is less than stellar, so their firms will have higher survival rates, even as they create less employment. On the other hand, we have ambitious entrepreneurs, with prior work experience at higher-performing larger firms. They start on a bigger scale, have more aggressive growth plans for their company, and need to take more risks to achieve higher growth, therefore face higher failure rates.

Choice to stay or leave an industry

Firms that display higher network concentration may be ones that remain in the same industry as the entering employees (hence the larger starting size and initial wage since industry specific productivity is more transparent). Individuals who jointly choose to move into being involved with a start-up may be more likely to do so within the same industry. Therefore, employees have industry specific skills and knowledge gathered from their previous place of employment and as such are more likely to survive. Firms that display low network concentration may reflect the founders' / employees' choice to switch into a different industry. This shift may reduce the odds of survival due to their lack of familiarity with the internal workings of the new industry. Conditional on surviving these firms with "industry switchers" catch up and do as well or better (steeper learning curve, leading to initial growth). This is another possible explanation for our results.

This section shows that there are multiple interpretations for the results found in this paper. While we do not causally identify any single mechanism that leads to entrepreneurial success, this exercise still documents a striking relationship between network concentration and new firm success. We also show how sensitive results are to the definition of entrepreneurial success and the timeline over which our hypotheses are evaluated. Our results combined with these various interpretations gives the reader some direction in thinking through what successful entrepreneurship means.

3.4.4 Concerns with the data and estimation strategies

Unobserved heterogeneity and selection bias

This paper does not seek to make any causal claims. Firstly, we have yet to address the fact that there is heterogeneity in the number of previously conjoined workers in new firms. Unobserved heterogeneity can come from any factor that correlates with the the type of network links we observe, but that doesn't actually pass through the network. For example, firms with higher human capital could have higher social capital and this may result in over-emphasizing the impact of strong links. Secondly, the process by which the future entrepreneur selects which (if any) employees to hire away from the parent firm introduces selection bias: even if worker quality in the parent firm varies randomly, an entrepreneur that happens to have higher quality coworkers will enlist more of them than his unlucky counterpart, so what we observe as a higher preponderance of cluster networks in well-performing start-ups may in fact be due to the higher quality of workers. We do control for worker skill as much as possible (education, age, formal sector experience, former wage), but there are always residual unobserved characteristics, so the risk remains that we are over-estimating the impact of strong links (cluster networks) on firm performance. Thirdly, it is possible that the very quality of one's peers induces them into jointly forming a new firm.

A potential identification strategy is to limit the analysis to new firms with workers that have prior experience, but instrument for the length of that joint experience with industry shocks.

Untrackable employees

Employees for whom we have no job history may be new to the job market, or they may have been exclusively employed at informal firms in the past (since 1986 when our dataset begins). We have no measure of their past work experience, or whether they have established prior working relationships with other untrackable employees in the new firm. However, evidence from 2- and 3-employee firms suggests that untracked workers typically do not share common work experience, or this past informal joint work is not as valuable as common experience in the formal sector. Having said that, this observation brings up the question of what the role of the informal sector is in entrepreneurial activity.

Access to financing

Unfortunately, we have no data on firm financing or capital holdings. One can imagine that differential access to financing might be partially driving our results: if a firm has several founders with prior work experience, they will trust each other more and may pool their resources and/or apply together (and with better chances of success) for a bank loan. One way we can test the role of financing is to estimate the impact of employee networks on survival for industries with different fixed costs of entry. Should we find that in sectors with negligible returns to scale employee networks don't matter, this will cast a shadow of doubt on the overall result.

Divestitures from parent firms

Some new firms may in fact be divestitures from old firms. These divestitures will have several advantages over true new firms, including accumulated capital stock, brand name, established supplier relations, etc. They will also typically have high employee network concentration, so they may be biasing our results upward. We plan to isolate divestitures by the share of employees of the parent establishment (plant/office branch) that is accounted for by employees transferred to the new firm.

More specifically, we propose to identify a new firm as a divestiture if 70% or more of the parent establishment's employees transferred to the new firm. In addition, we can use the legal form of the new entrant to identify divestitures[22].

New firm buy-outs

Some new firms may drop out of the sample because they are bought by larger corporations. If we don't account for this possibility, we will incorrectly label some well-performing firms as bad performers, especially since we use survival as our measure of success. However, since these buy-outs are likely to account for only a minority of firm disappearances, we believe that failing to account for them only biases our results towards zero.

In the future, we will attempt to identify buy-outs by the share of the new firm employees that can be later found working together at their next job, employing a rule analogous to the proposed rule for identifying divestitures: if 70% or more of a new firm's employees (call this firm *A*) can be found employed together at a previously existing firm *B* after the disappearance of firm *A*, then eliminate firm *A* from the sample as a likely buy-out.

3.5 Conclusion

This paper provides some strong empirical evidence for multiple phenomena in entrepreneurship. Our main result is that the strength of ties do matter for entrepreneurial success. Specifically, firms whose employees have previous joint work experience tend to survive longer, are bigger at inception and have a higher initial average wages. Conversely, joint previous work experience is *negatively* associated with new firm *growth* up to three years after entry, and uncorrelated with growth thereafter. However, this latter effect is of smaller magnitude, so that the initial performance differentials persist, and firms with higher initial network concentrations have more employees and higher wages up to six years after entry - the maximum length of time we can track them.

[22]see Hirakawa, Muendler and Rauch (December 2010)

We also find that new firms wherein the initial employees were previously employed at smaller firms are more likely to survive than new firms whose employees have larger firm experience instead. Despite this, employees from larger parent firms are positively associated with success as defined by a new firm's initial size, wage and employment growth rates.

As incidental findings, we observe that firm success is positively correlated with mean employee education and negatively correlated with mean employee age. Our final observation is that new firms that have a larger proportion of "untracked" employees, i.e. employees newly entering the labor force or transitioning in from the informal sector, are less likely to succeed.

The literature on entrepreneurial networks is a fairly thin one, and this paper provides at least a sliver of insight into the types of links that contribute to entrepreneurial success. While we use the most basic of measures to estimate network ties, we are able to point to startling and significant trends that have yet to be addressed in the empirical literature. In pointing to these trends this paper guides readers to thinking more carefully about what being a successful entrepreneur means, how this should be measured and what characteristics lead to successful outcomes.

From a policy perspective, thinking about these issues should provide some insight into why individuals start new firms, whom they choose as business partners, and which ventures we should encourage when trying to promote truly entrepreneurial activity.

Acknowledgements

We thank Roger Gordon, Gordon Hanson, Sam Bazzi, Silke Forbes, Craig McIntosh, Philip Neary and Chris Woodruff for many helpful comments. We also thank Marc Muendler for providing us with access to the RAIS data.

3.6 Figures

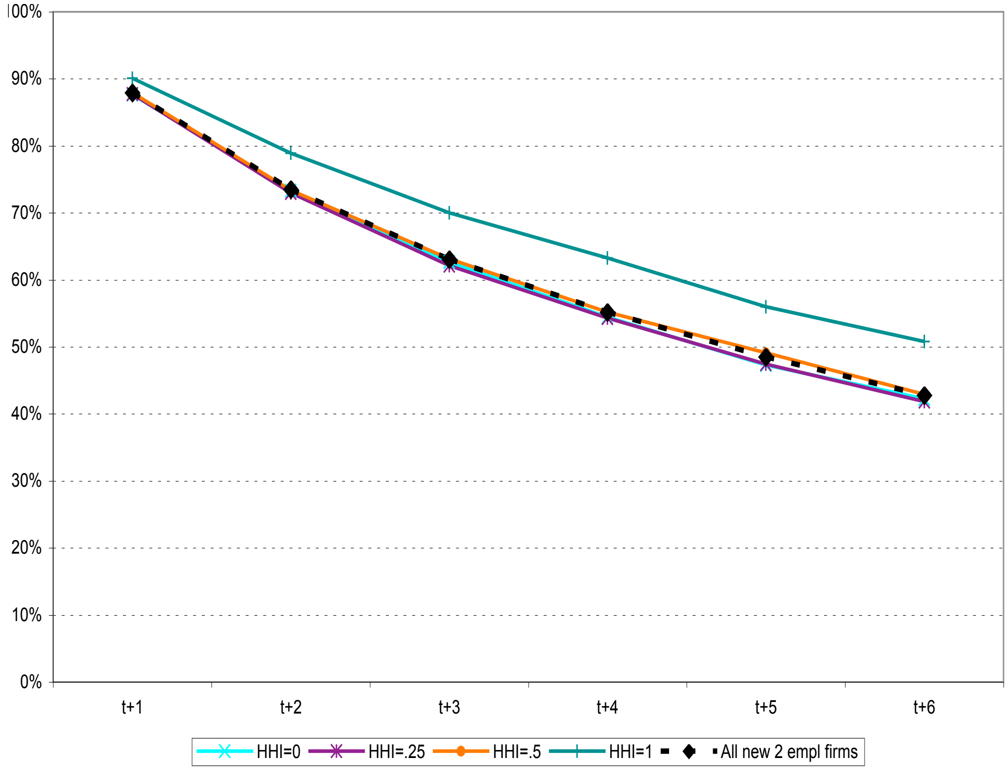

Figure 3.1: Average survival rates for 2-employee firms, by HHI_1

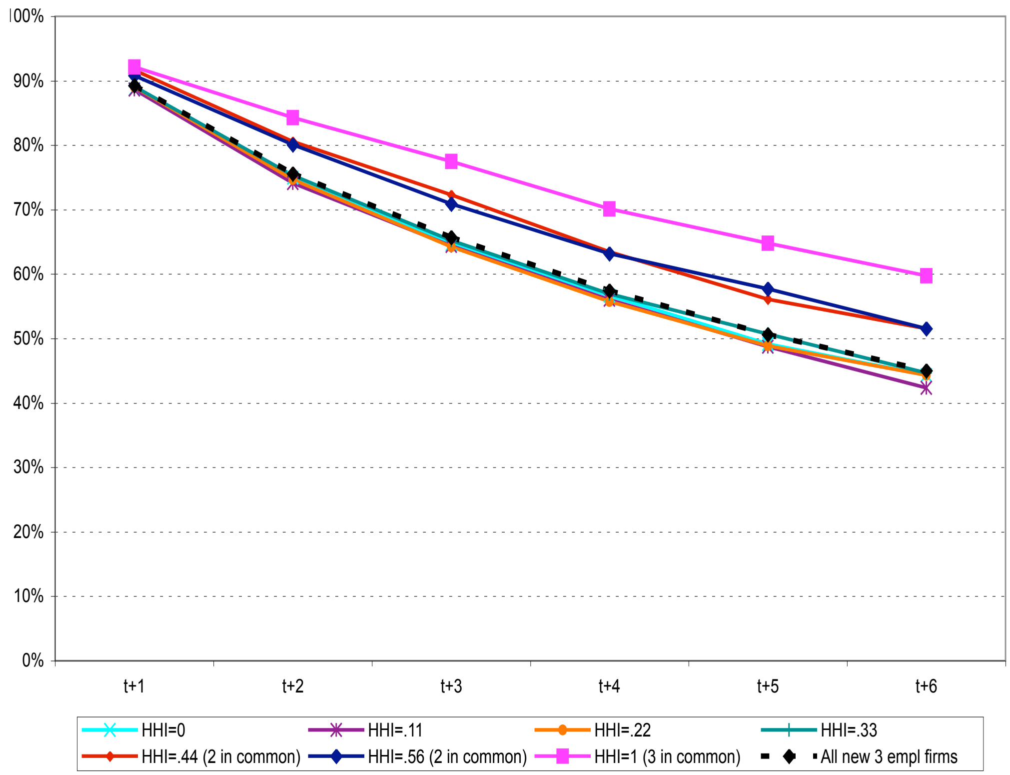

Figure 3.2: Average survival rates for 3-employee firms, by HHI_1

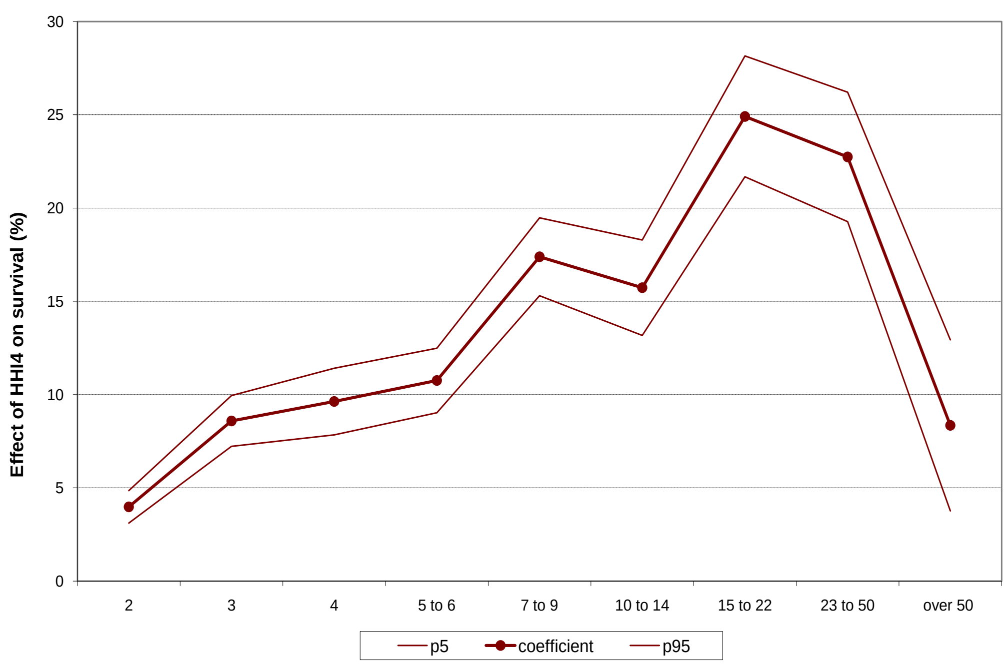

Figure 3.3: Survival impact of HHI_4 at t+3, by initial firm size

3.7 Tables

Table 3.1: HHI values in 2-employee firms

HHI_1	HHI_2	HHI_3	HHI_4	
0	.	0.5	0	No worker can be tracked
0.25	.	0.5	0	One can be tracked
0.5	0.5	0.5	0	Both tracked, but to different parents
1	1	1	1	Both tracked to same parent

Table 3.2: HHI values in 3-employee firms

HHI_1	HHI_2	HHI_3	HHI_4	
0	.	0.33	0	No worker can be tracked
0.11	.	0.33	0	One can be tracked
0.22	0.5	0.33	0	Two can be tracked, to different parents
0.33	0.33	0.33	0	All three can be tracked, all to different parents
0.44	1	0.56	0.33	Two can be tracked, to the same parent
0.56	0.56	0.56	0.33	All three can be tracked, 2 to the same parent
1	1	1	1	All three can be tracked tracked, all to same parent

Table 3.3: Average parent size breakdown

Unknown parent size	13.8%
Small parents (less than 10)	16.3%
Medium parents (10 to 100)	37.1%
Large parents (100 to 1,000)	25.7%
Very large parents (1,001+)	7.2%

Table 3.4: HHI_3 and HHI_4, by initial number of employees

Initial size	N	Percent of total	HHI_3 mean	HHI_3 sd	HHI_4 mean	HHI_4 sd
2	239,078	32.4%	0.530	0.119	0.060	0.237
3	128,716	17.5%	0.379	0.135	0.068	0.203
4	81,014	11.0%	0.306	0.146	0.075	0.194
5	55,582	7.5%	0.263	0.149	0.079	0.186
6	39,293	5.3%	0.234	0.151	0.081	0.182
7	29,201	4.0%	0.216	0.154	0.085	0.180
8	22,500	3.1%	0.198	0.151	0.084	0.173
9	17,864	2.4%	0.185	0.153	0.084	0.172
10	14,576	2.0%	0.178	0.156	0.086	0.174
11	11,685	1.6%	0.168	0.156	0.085	0.171
12	9,542	1.3%	0.161	0.154	0.085	0.168
13	8,161	1.1%	0.162	0.164	0.092	0.178
14	6,993	0.9%	0.161	0.169	0.097	0.182
15	6,007	0.8%	0.154	0.165	0.094	0.177
Subtotal	670,212	91.0%	0.369	0.192	0.071	0.208
Total	736,838	100.0%	0.349	0.202	0.075	0.207

Table 3.5: HHI_3 and HHI_4, by bins of initial number of employees

Initial size bins	N	percent of total	HHI_3 mean	HHI_3 sd	HHI_4 mean	HHI_4 sd
2	239,078	32.4%	0.530	0.119	0.060	0.237
3-4	209,730	28.5%	0.351	0.144	0.071	0.199
5-10	179,016	24.3%	0.226	0.155	0.082	0.180
11-20	62,453	8.5%	0.157	0.162	0.091	0.175
21-50	32,328	4.4%	0.134	0.174	0.103	0.180
51-100	8,602	1.2%	0.133	0.188	0.120	0.191
101+	5,631	0.8%	0.203	0.262	0.198	0.264
Total	736,838	100.0%	0.349	0.202	0.075	0.207

Table 3.6: Survival regressions for HHI_4

Survival at:	t+1	t+2	t+3	t+4	t+5	t+6
	(1)	(2)	(3)	(4)	(5)	(6)
HHI4	2.42	6.55	8.59	9.30	8.72	8.83
	(.17)***	(.27)***	(.34)***	(.43)***	(.57)***	(.85)***
Share trackable	1.32	1.81	2.23	2.23	3.25	3.37
	(.19)***	(.28)***	(.35)***	(.42)***	(.55)***	(.76)***
Mean employee age	-.06	-.06	-.05	-.06	-.06	-.10
	(.007)***	(.01)***	(.01)***	(.02)***	(.02)*	(.03)**
Mean years schooling	.30	.28	.22	.16	.12	.14
	(.02)***	(.02)***	(.03)***	(.03)***	(.04)*	(.06)
Mean previous wage	.07	.14	.16	.16	.13	.05
	(.01)***	(.02)***	(.02)***	(.02)***	(.03)***	(.04)
Unknown parent size	-1.91	-2.68	-2.92	-2.88	-3.24	-3.56
	(.21)***	(.32)***	(.40)***	(.48)***	(.63)***	(.85)***
Medium parents	-.47	-.68	-.86	-.92	-.96	-1.55
(10 to 100)	(.11)***	(.17)***	(.22)***	(.27)**	(.36)*	(.50)*
Large parents	-1.71	-2.65	-2.91	-2.84	-2.96	-3.44
(100 to 1,000)	(.12)***	(.19)***	(.24)***	(.29)***	(.38)***	(.53)***
Very large parents	-1.94	-2.93	-3.12	-2.89	-2.96	-3.21
(1,001+)	(.17)***	(.25)***	(.30)***	(.37)***	(.47)***	(.67)***
Initial size categories	Yes	Yes	Yes	Yes	Yes	Yes
Cohorts	Yes	Yes	Yes	Yes	Yes	Yes
Sectors	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	658,220	538,157	421,881	310,560	186,284	93,266
R^2	.02	.03	.04	.05	.06	.07

Notes: Dependent variable is the survival indicator multiplied by 100. Robust standard errors are shown in parentheses. Significance indicated is at 1%(*), 0.1%(**), and 0.01%(***).

Table 3.7: Survival regressions for HHI_4, adding interaction with initial firm size

Survival at:	t+1	t+2	t+3	t+4	t+5	t+6
	(1)	(2)	(3)	(4)	(5)	(6)
HHI4	1.84	4.76	4.70	3.78	1.78	2.40
	(.29)***	(.44)***	(.57)***	(.70)***	(.96)	(1.46)
Ln(initial firm size)	1.85	3.21	3.06	2.65	2.06	2.13
	(.05)***	(.08)***	(.10)***	(.12)***	(.16)***	(.22)***
HHI4*Ln(initial size)	.47	1.41	3.06	4.34	5.51	5.15
	(.16)*	(.25)***	(.33)***	(.42)***	(.60)***	(.95)***
Share trackable	.98	1.18	1.38	1.46	2.53	2.76
	(.18)***	(.28)***	(.35)***	(.42)**	(.54)***	(.76)**
Mean employee age	-.06	-.06	-.05	-.06	-.07	-.10
	(.007)***	(.01)***	(.01)***	(.02)***	(.02)*	(.03)**
Mean years schooling	.30	.28	.22	.16	.11	.13
	(.02)***	(.02)***	(.03)***	(.03)***	(.04)	(.06)
Mean previous wage	.06	.13	.15	.16	.13	.05
	(.01)***	(.02)***	(.02)***	(.02)***	(.03)***	(.04)
Unknown parent size	-2.11	-3.03	-3.35	-3.23	-3.53	-3.76
	(.21)***	(.32)***	(.40)***	(.48)***	(.63)***	(.85)***
Medium parents	-.38	-.51	-.64	-.74	-.80	-1.40
(10 to 100)	(.11)**	(.17)*	(.22)*	(.27)*	(.35)	(.50)*
Large parents	-1.69	-2.60	-2.81	-2.75	-2.81	-3.31
(100 to 1,000)	(.12)***	(.19)***	(.24)***	(.29)***	(.38)***	(.53)***
Very large parents	-2.03	-3.11	-3.34	-3.13	-3.23	-3.46
(1,001+)	(.17)***	(.25)***	(.31)***	(.37)***	(.48)***	(.67)***
Cohorts	Yes	Yes	Yes	Yes	Yes	Yes
Sectors	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	658,220	538,157	421,881	310,560	186,284	93,266
R^2	.02	.03	.04	.05	.06	.07

Notes: Dependent variable is the survival indicator multiplied by 100. Robust standard errors are shown in parentheses. Significance indicated is at 1%(*), 0.1%(**), and 0.01%(***).

Table 3.8: Performance at entry and conditional on survival, for HHI_4

	t		t+3		t+6	
	ln(empl)	ln(wage)	ln(empl)	ln(wage)	ln(empl)	ln(wage)
	(1)	(2)	(3)	(4)	(5)	(6)
HHI4	.60	.10	.39	.06	.33	.03
	(.006)***	(.003)***	(.01)***	(.005)***	(.03)***	(.01)***
Share trackable	-.33	.35	-.04	.32	-.02	.30
	(.004)***	(.003)***	(.01)***	(.005)***	(.03)	(.01)***
Mean employee age	.0005	.006	-.002	.002	-.002	.0001
	(.0002)***	(.0001)***	(.0004)***	(.0002)***	(.0009)**	(.0004)
Mean years schooling	-.005	.02	.007	.02	.01	.01
	(.0004)***	(.0003)***	(.0009)***	(.0004)***	(.002)***	(.001)***
Mean previous wage	.007	.04	.02	.04	.02	.03
	(.0003)***	(.0005)***	(.0007)***	(.0007)***	(.002)***	(.002)***
Unknown parent size	-.19	-.003	-.11	-.005	-.09	-.04
	(.003)***	(.003)	(.009)***	(.005)	(.03)***	(.01)***
Medium parents	.47	.06	.43	.06	.37	.05
(10 to 100)	(.002)***	(.002)***	(.005)***	(.003)***	(.01)***	(.007)***
Large parents	.57	.07	.54	.07	.50	.05
(100 to 1,000)	(.003)***	(.002)***	(.006)***	(.003)***	(.02)***	(.007)***
Very large parents	.09	.006	.15	.02	.17	.02
(1,001+)	(.004)***	(.003)**	(.009)***	(.004)***	(.02)***	(.01)*
Cohorts	Yes	Yes	Yes	Yes	Yes	Yes
Sectors	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	619,173	619,173	246,248	246,248	39,454	39,454
R^2	.18	.28	.15	.25	.16	.25

Notes: Robust standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***). Number of employees and average wages reported are for workers employed on December 31st of each year. Columns 1 and 2 include all firms born between 1995 and 2000, but since some no longer have employees by the end of their first year, the number of observations is lower than in column 1 of tables 3.6 and 3.7. Columns 3 and 4 include firms born in 1995 through 1998, and which have survived through their third year. Columns 5 and 6 include firms born in 1995, and which have survived through 2001.

Table 3.9: Performance at entry and growth conditional on survival, for HHI_4

	performance at entry		growth from t to t+3		growth from t+3 to t+6	
	ln(empl)	ln(wage)	ln(empl)	ln(wage)	ln(empl)	ln(wage)
	(1)	(2)	(3)	(4)	(5)	(6)
HHI4	.55	.09	-.12	-.03	.01	-.001
	(.02)***	(.01)***	(.02)***	(.01)***	(.02)	(.009)
Share trackable	-.29	.40	.14	-.07	.02	-.05
	(.02)***	(.01)***	(.02)***	(.01)***	(.02)	(.009)***
Mean employee age	.002	.004	-.002	-.004	-.003	-.0007
	(.0007)***	(.0005)***	(.0007)**	(.0004)***	(.0007)***	(.0003)**
Mean years schooling	-.001	.02	.008	-.007	.0001	-.0002
	(.002)	(.001)***	(.002)***	(.0009)***	(.002)	(.0008)
Mean previous wage	.01	.05	.009	-.006	.002	-.004
	(.002)***	(.002)***	(.002)***	(.0008)***	(.001)	(.0006)***
Unknown parent size	-.18	-.03	.03	-.02	-.006	-.005
	(.02)***	(.01)**	(.02)	(.01)	(.02)	(.01)
Medium parents	.46	.08	.07	.002	-.01	-.005
(10 to 100)	(.01)***	(.008)***	(.01)***	(.007)	(.01)	(.006)
Large parents	.59	.08	.10	.00009	.007	-.005
(100 to 1,000)	(.01)***	(.009)***	(.01)***	(.007)	(.01)	(.006)
Very large parents	.15	.04	.09	.0006	-.01	-.02
(1,001+)	(.02)***	(.01)***	(.02)***	(.009)	(.02)	(.008)**
Initial size categories			Yes	Yes	Yes	Yes
Sectors	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	35,864	35,864	35,864	35,864	35,864	35,864
R^2	.21	.32	.1	.04	.04	.03

Notes: Robust standard errors are shown in parentheses. Significance indicated is at 10%(*), 5%(**), and 1%(***). Number of employees and average wages reported are for workers employed on December 31st of each year. Only firms with end-of-year employees at t, t+3, and t+6 are included.

3.8 Appendix

We show survival regression results for the alternate Herfindahl-Hirschman Index variants (HHI_1, HHI_2, and HHI_3).

First, we compare these three indices with each other (since HHI_4, our baseline measure, is just a rescaling of HHI_3). Note that as we move from table 3.10 to table 3.12, the coefficients on the HHI_1 and HHI_3 index are remarkably similar, given the different definitions. The impact of "share trackable" does change in the direction we expect, since HHI_1 unfairly gives zero share to employees who are new to the formal labor market.

Comparing tables 3.11 and 3.12, we find that the coefficients are once again similar. The number of observations drops, since HHI_2 is only defined for firms with two or more trackable workers.

When we compare the baseline results from table 3.6 to the corresponding results in table 3.12 (that uses HHI_3), we observe that the coefficients on the HHI variable are smaller in the former; this is natural, since HHI_4 takes on a wider range of values for small firms (0 to 1, instead of 1/N to 1 for HHI_3), and small firms form the majority of new entrants. Other coefficients are largely unchanged.

Table 3.10: Survival regressions for HHI_1

Survival at:	t+1	t+2	t+3	t+4	t+5	t+6
	(1)	(2)	(3)	(4)	(5)	(6)
HHI1	2.90	8.57	11.95	12.94	12.77	12.73
	(.24)***	(.37)***	(.49)***	(.61)***	(.81)***	(1.22)***
Share trackable	.56	-.62	-1.32	-1.61	-.70	-.53
	(.22)	(.33)	(.42)*	(.51)*	(.67)	(.95)
Mean employee age	-.06	-.06	-.05	-.06	-.06	-.10
	(.007)***	(.01)***	(.01)***	(.02)***	(.02)*	(.03)**
Mean years schooling	.30	.28	.22	.17	.12	.14
	(.02)***	(.02)***	(.03)***	(.03)***	(.04)*	(.06)
Mean previous wage	.07	.14	.16	.17	.14	.06
	(.01)***	(.02)***	(.02)***	(.02)***	(.03)***	(.04)
Unknown parent size	-1.87	-2.57	-2.77	-2.72	-3.09	-3.41
	(.21)***	(.32)***	(.40)***	(.48)***	(.63)***	(.85)***
Medium parents	-.50	-.73	-.89	-.97	-.98	-1.57
(10 to 100)	(.11)***	(.17)***	(.22)***	(.27)**	(.36)*	(.50)*
Large parents	-1.74	-2.69	-2.92	-2.86	-2.93	-3.43
(100 to 1,000)	(.12)***	(.19)***	(.24)***	(.29)***	(.38)***	(.53)***
Very large parents	-1.96	-2.99	-3.19	-2.97	-3.04	-3.28
(1,001+)	(.17)***	(.25)***	(.31)***	(.37)***	(.47)***	(.67)***
Initial size categories	Yes	Yes	Yes	Yes	Yes	Yes
Cohorts	Yes	Yes	Yes	Yes	Yes	Yes
Sectors	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	658,220	538,157	421,881	310,560	186,284	93,266
R^2	.02	.03	.04	.05	.06	.07

Notes: Dependent variable is the survival indicator multiplied by 100. Robust standard errors are shown in parentheses. Significance indicated is at 1%(*), 0.1%(**), and 0.01%(***).

Table 3.11: Survival regressions for HHI_2

Survival at:	t+1	t+2	t+3	t+4	t+5	t+6
	(1)	(2)	(3)	(4)	(5)	(6)
HHI2	3.45	8.59	11.52	12.63	12.75	13.09
	(.20)***	(.31)***	(.41)***	(.52)***	(.69)***	(1.03)***
Share trackable	3.38	6.75	8.63	9.51	10.63	10.46
	(.24)***	(.37)***	(.47)***	(.58)***	(.77)***	(1.09)***
Mean employee age	-.07	-.08	-.06	-.07	-.07	-.10
	(.008)***	(.01)***	(.01)***	(.02)**	(.02)*	(.03)*
Mean years schooling	.31	.31	.27	.20	.16	.20
	(.02)***	(.03)***	(.03)***	(.04)***	(.05)*	(.07)*
Mean previous wage	.09	.20	.22	.22	.17	.08
	(.01)***	(.02)***	(.02)***	(.03)***	(.04)***	(.06)
Unknown parent size	-3.37	-4.54	-5.07	-5.83	-5.54	-3.45
	(.39)***	(.58)***	(.72)***	(.88)***	(1.15)***	(1.51)
Medium parents	-.37	-.76	-.91	-1.08	-.93	-1.54
(10 to 100)	(.12)*	(.20)**	(.26)**	(.32)**	(.43)	(.61)
Large parents	-1.87	-2.97	-3.13	-3.18	-3.02	-3.48
(100 to 1,000)	(.14)***	(.22)***	(.28)***	(.34)***	(.45)***	(.64)***
Very large parents	-2.29	-3.47	-3.66	-3.54	-2.89	-3.25
(1,001+)	(.20)***	(.30)***	(.37)***	(.45)***	(.58)***	(.83)***
Initial size categories	Yes	Yes	Yes	Yes	Yes	Yes
Cohorts	Yes	Yes	Yes	Yes	Yes	Yes
Sectors	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	518,479	423,251	330,567	243,080	146,027	72,723
R^2	.02	.04	.05	.06	.07	.08

Notes: Dependent variable is the survival indicator multiplied by 100. Robust standard errors are shown in parentheses. Significance indicated is at 1%(*), 0.1%(**), and 0.01%(***).

Table 3.12: Survival regressions for HHI_3

Survival at:	t+1	t+2	t+3	t+4	t+5	t+6
	(1)	(2)	(3)	(4)	(5)	(6)
HHI3	3.34	9.28	12.87	14.09	14.00	14.29
	(.24)***	(.37)***	(.48)***	(.61)***	(.82)***	(1.23)***
Share trackable	1.33	1.78	2.03	2.00	2.89	3.01
	(.19)***	(.28)***	(.35)***	(.42)***	(.55)***	(.77)***
Mean employee age	-.06	-.06	-.05	-.06	-.06	-.10
	(.007)***	(.01)***	(.01)***	(.02)***	(.02)*	(.03)**
Mean years schooling	.30	.28	.22	.16	.12	.13
	(.02)***	(.02)***	(.03)***	(.03)***	(.04)*	(.06)
Mean previous wage	.06	.14	.15	.16	.13	.05
	(.01)***	(.02)***	(.02)***	(.02)***	(.03)***	(.04)
Unknown parent size	-1.93	-2.73	-2.99	-2.95	-3.31	-3.62
	(.21)***	(.32)***	(.40)***	(.48)***	(.63)***	(.85)***
Medium parents	-.47	-.67	-.81	-.86	-.87	-1.44
(10 to 100)	(.11)***	(.17)***	(.22)**	(.27)*	(.36)	(.51)*
Large parents	-1.70	-2.62	-2.82	-2.74	-2.80	-3.28
(100 to 1,000)	(.12)***	(.19)***	(.24)***	(.29)***	(.38)***	(.53)***
Very large parents	-1.96	-2.98	-3.18	-2.96	-3.03	-3.26
(1,001+)	(.17)***	(.25)***	(.30)***	(.37)***	(.47)***	(.67)***
Initial size categories	Yes	Yes	Yes	Yes	Yes	Yes
Cohorts	Yes	Yes	Yes	Yes	Yes	Yes
Sectors	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	658,220	538,157	421,881	310,560	186,284	93,266
R^2	.02	.03	.04	.05	.06	.07

Notes: Dependent variable is the survival indicator multiplied by 100. Robust standard errors are shown in parentheses. Significance indicated is at 1%(*), 0.1%(**), and 0.01%(***).

3.9 Bibliography

Aldrich, H., *Organizations evolving*, Sage Pubns, 1999.

Elfring, Tom and Willem Hulsink, "Networks in Entrepreneurship: The Case of High-technology Firms," *Small Business Economics*, 2003, *21*, 409–422. 10.1023/A:1026180418357.

Giannetti, M. and A. Simonov, "Social Interactions and Entrepreneurial Activity," *Journal of Economics and Management Strategy*, 2009.

Gompers, P., J. Lerner, and D. Scharfstein, "Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986 to 1999," *Journal of Finance*, 2005, pp. 577–614.

Granovetter, M., "The strength of weak ties: A network theory revisited," *Sociological theory*, 1983, pp. 201–233.

Hirakawa, O., M.A. Muendler, and J.E. Rauch, "Employee spinoffs and other entrants: stylized facts from Brazil," *IGC Working Paper, 10/0879.*, December 2010.

Hvide, H.K., "Firm size and the quality of entrepreneurs," *CEPR Discussion Paper No. 4979*, 2005.

Lazear, Edward P., "Balanced Skills and Entrepreneurship," *The American Economic Review*, 2004, *94* (2), pp. 208–211.

Lerner, Josh and Ulrike Malmendier, "With a Little Help from My (Random) Friends: Success and Failure in Post-Business School Entrepreneurship," Working Paper 16918, National Bureau of Economic Research March 2011.

Nanda, R. and J. Sorensen, "Peer effects and entrepreneurship," *Harvard Business School Entrepreneurial Management Working Paper No. 08-051*, 2009.

Rauch, J.E. and J. Watson, "Clusters and bridges in networks of entrepreneurs," *The Missing Links: Formation and Decay of Economic Networks*, 2007, pp. 210–235.

Shane, S. and S.A. Shane, *A general theory of entrepreneurship: The individual-opportunity nexus*, Edward Elgar Pub, 2004.

Chapter 4

Employee Spinoffs and Other Entrants: Stylized Facts from Brazil

Abstract

We gauge the prevalence and performance of firms founded as employee spinoffs, relative to new firms without parents and to diversification ventures of existing firms entering new industries. Using a comprehensive linked employer-employee database from Brazil for the period 1995-2001, we are able to identify an employee spinoff either when the director/manager moved from a parent in the same industry or when one-quarter of the employees shifted from a common parent. Depending on definition, employee spinoffs account for between one-sixth and one-third of the new firms in Brazil's private sector during this period. Regardless of definition, size at entry is larger for employee spinoffs than for new firms without parents but smaller than for diversification ventures of existing firms. Similarly, survival rates for employee spinoffs are higher than for new firms without parents and comparable to those for diversification ventures of existing firms. These results suggest that we can think of some part of a firm's productivity and riskiness as embodied in the firm's employees and portable by them to a new firm.

4.1 Introduction

Where do new firms come from? One answer is from other firms: firms lose employees, who spin off to form their own businesses. A burgeoning literature seeks to quantify the employee spinoff phenomenon. We use a comprehensive linked employer-employee data set for Brazil to count spinoffs using precise and replicable criteria and compare basic indicators of their performance to those of other entrants.

Klepper and Sleeper (2005) count spinoffs, their "parents," and other new entrants in the U.S. laser industry through 1994. Franco and Filson (2006) conduct a similar study for the rigid disk drive industry in the period 1977-1997. Eriksson and Kuhn (2006) compare the entry and survival of spinoffs with sizes from two to ten employees to other new small firms, using a linked employer-employee data set for

Danish private-sector firms during the period 1981-2000. Our sampling approach is closest to that of Eriksson and Kuhn (2006), but we include new firms with more than ten employees and existing firms entering new industries in our analysis. We follow the classic work of Dunne et al. (1988) and examine firm entry size and exit rates relative to both new firms that are not spinoffs and existing firms entering new industries. Dunne et al. used U.S. data and did not distinguish spinoffs from other new firms. We use Brazilian data because they allow identification of employee spinoffs, including distinguishing them from employer-initiated divestitures. The Global Entrepreneurship Monitor consistently ranks Brazil, a large and diversified economy and a leading emerging market, among the most entrepreneurial economies in the world measured by the prevalence of nascent and new firms in the economy (Reynolds et al., 2000; Minniti et al., 2005).

Our results for Brazil during the period 1995-2001 are that, depending on spinoff definition, employee spinoffs account for around one-sixth of new firms with salaried management and for one-third of new firms with five or more employees—excluding those new firms with state ownership, cooperatives, any type of holding company, and foreign subsidiaries. Regardless of spinoff definition, size at entry for employee spinoffs is larger than for new firms without parents but smaller than for employer-initiated divestitures or diversification ventures of existing firms. Similarly, exit rates for employee spinoffs are lower than for new firms without parents but greater than for employer-initiated divestitures and comparable to those for diversification ventures of existing firms. Across the four types of entrants studied (unrelated new firms, spinoffs, divestitures, and diversification ventures), size at entry is largest for diversification ventures and exit rates are lowest for divestitures.

One way to interpret these findings regarding performance is to consider the four entrant types as embodying differing levels of initial capital endowments (human, organizational, and other forms) and differing levels of insurance against idiosyncratic risk or uncertainty about product success. Unrelated new firms have the least amount of both capital and insurance. Spinoffs may draw some capital from their parent firms, but face the same uncertainties as an unrelated new firm. Diversification ventures have the capital of the parent firms and some of their lower risk, but face some uncertainty about product success. Finally, divestitures draw upon the capital of their parent firms when starting up and generally do not face any of the product uncertainty of a diversification. Inheritance from a parent firm is valuable, according to the model of Jovanovic (1982), because the parent has been selected for high productivity relative to the typical new firm by virtue of having survived for some period of time. The participation of employee spinoffs in this inheritance despite lack of management by the parent firm shows that some part of the parent's productivity and riskiness is embodied in its employees and portable by them to a new firm.

Section 4.2 provides an overview of the literatures on employee spinoffs and divestitures. Section 4.3 describes the data source. Section 4.4 introduces our classification of entrants into employee spinoffs, divestitures, diversification ventures, and unrelated new firms, and section 4.5 compares the frequencies of these entrant types. Section 4.6 documents the performance of employee spinoffs relative to other entrants, and section 4.7 concludes. The Appendix provides details on the data source and the empirical implementation of definitions.

4.2 Related Literature

The earliest papers on employee spinoffs were motivated by high-profile examples in the U.S. high-tech sector. Subsequently the literature broadened beyond this narrow focus. After all, standard problems with eliciting effort inside organizations (Alchian and Demsetz, 1972; Holmstrom, 1982) can motivate employee spinoffs in any sector of the economy. The literature on high-tech spinoffs compares their performance to new plants of parents with the idea that new plants of existing firms exploit employees' innovations in house. Parents, or incumbents, may have some advantages of scale, scope, tax, or information that will allow them to commercialize a discovery made by employees more profitably than a new firm started by employees (Klepper, 2001). Complementary assets of incumbent firms such as production capabilities, sales channels and marketing capabilities may be crucial to bringing innovations to the market quickly and successfully (Teece, 1986). On the other hand, spinoffs are free from "organizational inertia" that incumbents might possess, and which can prevent incumbents from adjusting to a new environment (Hannan and Freeman, 1984; Henderson and Clark, 1990). New plants of incumbent firms are more likely to inherit established processes and routines of their parent firms, which may preclude them from acting quickly, especially when the industry they enter is rapidly changing. In this regard, the models of Anton and Yao (1995) and Wiggins (1995) suggest that the innovations commercialized by spinoffs are more likely to be path-breaking or to be opening new sub-markets. If that conjecture is correct, spinoffs may show greater success than new plants of incumbents.

As the more recent literature has broadened away from high-tech spinoffs so that possession of a technological innovation is not required to start a new firm, it has compared the performance of spinoffs to that of other new firms rather than to new plants of existing firms. Cabral and Wang (2008) have a model and evidence from the automobile industry showing that spinoffs from surviving firms are superior to other new firms because the spinoffs are self-selected from all employees for entrepreneurial talent, whereas spinoffs from dying firms are negatively selected (at least relative to spinoffs from surviving firms) because all employees are looking to "jump ship" regardless of entrepreneurial ability. The findings of Eriksson and Kuhn (2006) support Cabral and Wang in that spinoffs from surviving firms have lower exit risk than spinoffs from dying firms, which in turn have equal or lower exit risk than other new firms. Hvide (2005) argued and presented evidence that spinoffs from large firms should be positively selected relative to spinoffs from small firms, because small firms can accurately recognize and reward employee ideas whereas large firms can only offer a higher wage, leading employees with the best ideas to leave and start their own firms.

A separate literature analyzes divestitures and corporate spinoffs. In contrast to an employee-initiated spinoff, a divestiture is a management-initiated new firm.

Common forms of divestitures are corporate spinoffs into standalone firms, or new firms that emerge as the results of parent firms' mergers and acquisitions, or new firms from a splitup of the parent firm into separate companies through equity transfers. A branch of the divestiture literature examines performance. Cusatis et al. (1993) document that, in addition to abnormal positive stock returns for the parent firm on the divestiture announcement date, both divestitures and their parents experience significantly positive abnormal returns for up to three years after announcement. Nandy and Chemmanur (2005) use large U.S. plant panel data, combined with stock return data for their firms, and document that productivity improves at the parents' plants and, to a lesser degree, at the divested plants upon divestiture, compared to plants at firms with no divestiture. We will see in Sections 4.4 through 4.6 below that distinguishing between divestitures and employee spinoffs is important in our data.

4.3 Data

We adopt a workforce-based definition of spinoffs and use employer-reported occupations. We study Brazilian data, where detailed occupational codes are available.[1] Our data derive from the linked employer-employee records RAIS (*Relação Anual de Informações Sociais* of the Brazilian labor ministry *MTE*), which offer comprehensive individual employee information on occupations, demographic characteristics and earnings, along with employer identifiers. By Brazilian law, every private or public-sector employer must report this information every year.[2] De Negri et al. (1998)

[1]To our knowledge, occupational information is currently neither available in the U.S. Longitudinal Business Database (LBD) nor in the Longitudinal Employer-Household Dynamics (LEHD) data base. In LEHD, educational information on the workforce is imputed by census tract. The reason for imputation is that U.S. unemployment insurance records, on which the employer-employee link is based state by state, do not typically offer educational information. But occupational information has not been imputed to date.

[2]RAIS primarily provides information to a federal wage supplement program (*Abono Salarial*), by which every employee with formal employment during the calendar year receives the equivalent of a monthly minimum wage. RAIS records are then shared across government agencies. An employer's failure to report complete workforce information can, in principle, result in fines proportional to the workforce size, but fines are rarely issued. In practice, employees and employers have strong incentives to ascertain complete RAIS records because payment of the annual public wage supplement is exclusively

compare labor force information in RAIS to that in a main Brazilian household survey (PNAD) and conclude that, when comparable, RAIS delivers qualitatively similar results to those in the national household survey. Menezes-Filho et al. (2008) apply the Abowd et al. (2001) earnings-estimation methodology to Brazil and show that labor-market outcomes from RAIS broadly resemble those in France and the United States, even after controlling for selection into formal-sector employment, except for unusually high returns to high school and college education and to experience among males. Appendix 4.9.1 presents further details on the data source.

A job observation in RAIS is identified by the employee ID, the employer's tax ID (CNPJ), and dates of job accession and separation. To avoid double-counting employees at new firms, we keep only one observation for each employer-employee pair, choosing the job with the earliest hiring date. If the employee has two jobs at the firm starting in the same month, we keep the highest paying one. The rules on tax ID assignments make it possible to identify new firms (the first eight digits of the tax ID) and new plants within firms (the last six digits of the tax ID). Appendix 4.9.2 discusses the relevant details on tax ID assignment. Our data include 71.1 million employees (with 556.3 million job spells) at 5.52 million plants in 3.75 million firms over the sixteen-year period 1986-2001 in any sector of the economy. We limit most attention to the years 1995-2001 and use the period 1986-1994 to define a *new firm* in 1995-2001 when its tax ID (first eight digits) appears for the first time. In addition, RAIS offers detailed industry information (at the four-digit *CNAE* level) starting in 1995. During this 7-year period, 1.54 million new firms and 2.17 million plants entered (of which 581 thousand new plants were created within incumbent firms). By 1995 macroeconomic stabilization had succeeded in Brazil. The Plano Real from August 1994 had brought inflation down to single-digit rates. Fernando Henrique Cardoso, who had enacted the Plano Real as Minister of Finance, became president, signalling a period of financial calm and fiscal austerity. Apart from a large exchange-rate devaluation in early 1999

based on RAIS. The ministry of labor estimates that well above 90 percent of all formally employed individuals in Brazil are covered in RAIS throughout the 1990s. Data collection is typically concluded by March following the year of observation.

and a subsequent switch from exchange-rate to inflation-targeting at the central bank, macroeconomic conditions remained relatively stable for the following years.

Occupational classifications in RAIS follow the *CBO* (*Classificação Brasileira de Ocupações*). This classification system with more than 350 categories allows us to identify management employees (directors or managers) for specific spinoff definitions. During our sample period, sectors are reported under the *CNAE* four-digit classification (*Classificação Nacional de Atividade Econômica*) for 654 industries, spanning all sectors of the economy. The level of detail is roughly comparable to the *NAICS 2007* five-digit level. RAIS reports an employee's earnings both as the monthly average wage during a year and as the December wage for jobs at year end. These earnings are measured in multiples of the current minimum wage, which we transform into Brazilian Real deflated to the August 1994 price level. Appendix 4.9.1 has further details on the earnings measures.

4.4 Definitions of Entrant Types

We take two complementary approaches to identifying employee spinoff firms in the RAIS data, and let each approach act as a check on the robustness of the other. In the first approach, we locate the human capital essential to founding the new firm in its director or manager.

Manager spinoff. *A director/manager spinoff is a new firm whose top paid director, or top paid manager if there are no directors, previously worked for an existing firm in the same 4-digit CNAE industry.*

The top paid director or manager may be the owner of the firm, or may have recruited financial backing from investors who own the firm but are not employed by it. Alternatively, investors may have recruited an experienced director or manager to run a new firm that was their idea. In the latter case, some (but not all) of the human capital essential to founding the new firm is embodied in the unobserved investors. Note that the manager spinoff definition will miss many "vertical" spinoffs, in which the top paid director or manager leaves his existing firm to independently produce an input he previously supplied to his former employer internally.[3] For example, an accountant for a manufacturing firm may start an accounting firm that caters to manufacturing industry. His new firm will not have the same 4-digit *CNAE* as his former employer and will therefore be missed by the manager spinoff definition.

Our second approach locates the human capital essential to founding the new firm in a group of employees that embodies its "core competence." Of course the core competence of a firm is unobserved, so we do not know which or how many employees embody its core competence. For help we turn to a fact about manager spinoffs: on average, the director or manager "brings along" from the parent 23 percent of the non-management employees of the new firm.[4] This suggests that a reasonable cutoff for the share of employees in the new firm that is needed to transfer essential technologies or work routines from the parent firm is one-quarter.[5]

Workforce spinoff. *A workforce spinoff is a new firm of five or more employees, at least 25 percent of whom previously worked for the same existing firm.*

We restrict this definition to new firms with five or more employees, because below five employees any new firm with an employee who can be traced to previous employment would automatically be a spinoff. In other words, by restricting ourselves to firms with five or more employees, we ensure that a "team" that embodies the core competence of the new firm must have at least two employees. An advantage of the workforce definition over the manager definition is that we are not restricted to firms with a paid director or manager, nor are we restricted to "horizontal" spinoffs. Moreover, the workforce definition can be implemented in linked employer-employee data sets in general, even if the data lack occupational information (as is the case in the United

[3]These vertical spinoffs are extensively documented for Taiwan in Chapter 7 of Shieh (1992).

[4]That is, on average 23 percent of the non-management employees of manager spinoffs, as counted in Table 4.2 below, are from the same parent firm as the top paid director or manager.

[5]Eriksson and Kuhn (2006) use one-half as the cutoff for defining a new firm with two to ten employees as a spinoff. However, they note that use of a 30 percent cutoff does not qualitatively change their findings.

States), so findings under this definition could be directly compared across countries. The obvious disadvantage is that without the presence of a director or manager it is entirely possible that no essential human capital is embodied in the group of employees.

Both spinoff definitions are vulnerable to the problem that the offspring firms may not be truly new. An existing firm that divests itself of one or more divisions creates a "new" firm that is likely to satisfy both of our spinoff definitions.[6] We receive some help with this problem from the coding of firms by *natureza juridica* (legal form) in the RAIS data set. By Brazilian commercial law, there are two broad categories of legal form: incorporated firms, and associations or partnerships without independent legal existence. Most important for our purposes, associations or partnerships cannot be owned by companies, but only by physical persons. So, if an employee spinoff is an association or partnership, it is not likely to be a divestiture (we call these "non-incorporated" legal forms). In contrast, spinoffs that are incorporated as Corporation under private control, Close corporation, or Limited liability company are quite possibly divestitures (we call these "incorporated" legal forms). Inverting the common criterion in the labor literature that a mass layoff is a reduction of the existing workforce by 30 percent or more (e.g. Jacobson et al., 1993), we label a new firm a divestiture if its *natureza juridica* is coded as Corporation under private control, Close corporation, or Limited liability company, or if it has unknown legal form, and if it absorbs 70 percent or more of the employees of a plant of an existing firm.[7]

Divestiture. *A divestiture is a new firm with natureza juridica coded as Corporation under private control, Close corporation, Limited liability company, or as unknown that absorbs 70 percent or more of the employees of a plant of an existing firm.*

[6]One might think the same problem could arise if a firm is sold, creating a "new" firm that is again likely to satisfy both of our spinoff definitions. However, as discussed in Appendix 4.9.2, a firm that is sold retains its firm identifier and therefore is not coded as a new firm in our data.

[7]We use the share of employees of an existing plant rather than an entire existing firm because a typical divestiture scenario is one in which a parent firm divests itself of a particular plant, which becomes a new firm. This conservative approach makes it more difficult to classify a new firm as an employee spinoff. Benedetto et al. (2007) use a cutoff of 80 percent of the employees of an existing firm shifting to another firm in order to cross-validate firm dynamics from administrative firm records with worker flow information. So as to check for the potential sensitivity of our later results to our choice of the cutoff at 70 percent, we control for the share of parent employees shifted in robustness regressions.

We exclude from our analysis branches of government, firms with state ownership, cooperatives, any type of holding company, and branches of foreign firms. In other words, we concentrate on Brazil's domestically-owned private sector. For our exhaustive and mutually exclusive classification of *natureza juridica* into non-incorporated legal forms, incorporated legal forms and inadmissible legal forms, see Table 4.8 in Appendix 4.9.3.

Table 4.9 in Appendix 4.9.4 summarizes the exhaustive and mutually exclusive classification of new firms resulting from these definitions. Appendix 4.9.4 also describes the classification procedure in more detail.

Our last entrant type is existing firms entering new industries.

Diversification venture. *A diversification venture is a group of one or more new plants within an existing firm in a different CNAE 4-digit industry than the existing firm.*

We follow Dunne et al. (1988) and do not consider new plants of an existing firm in the same industry as entrants.

4.5 Counting Employee Spinoffs, Other New Firms, and Diversification Ventures

Having defined types of entrants, we can assess their relative frequency in the Brazilian formal sector. The universe from which we sample consists of all new firms and diversification ventures with included legal form. Note that the pool of new firms from which manager spinoffs can be drawn is restricted to those with at least one director or manager, and the pool of new firms from which workforce spinoffs can be drawn is restricted to those with at least five employees. We therefore draw two samples from our universe, containing all new firms and diversification ventures with included legal form with at least one director or manager and with five or more employees, respectively. Table 4.1 shows that having at least five employees is more than four times more common in our universe than having at least one director or manager. Even the larger sample covers only 21.9 percent of our universe of new firms, though these account for 73.0 percent of employment. Coverage of diversification ventures is more than twice that of new firms in both samples, and the larger sample approaches complete coverage of employment for diversification ventures.[8]

Table 4.2 shows that manager spinoffs and workforce spinoffs respectively account for about one-sixth and nearly 30 percent of new firms in their samples. The ranking is to be expected given the greater restrictiveness of the manager spinoff definition. Under both definitions spinoffs account for larger shares of employees at time of entry than they do of counts of new firms. This holds even more strongly for divestitures. Differences in initial sizes across types of new firms and between new firms and diversification ventures will be examined in the next section.

We can assess the overlap between our two spinoff definitions by considering the subset of new firms with included legal form that have both a director/manager and at least five employees. There are 41,725 firms in this subset, of which 10,783 are manager spinoffs, 17,010 are workforce spinoffs, and 6,386 are both. Thus 59.2 percent of manager spinoffs are also workforce spinoffs but only 37.5 percent of workforce spinoffs are also manager spinoffs. This again emphasizes that the manager spinoff definition is more restrictive than the workforce spinoff definition.

Table 4.3 provides some perspective on the importance to the Brazilian economy of the new firms and ventures from which we sample. We examine the shares of these entrants in total Brazilian formal sector employment at the beginning, middle, and end of the period we cover. In any given year, the contributions to employment of the new firms and diversification ventures that enter in that year are in the neighborhood of four percent. At the end of the period, the contribution to employment of all the new firms and diversification ventures that entered and survived from 1995 to 2001 is 25.6 percent. This is despite the fact that these entrants exclude the public sector and foreign subsidiaries.

[8]Employment figures are the ones recorded in December of the new firms' or ventures' first year of appearance in RAIS. Averaging employment over the entire calendar year is too cumbersome in RAIS.

The bottom part of Table 4.3 reports the same figures for the larger and more representative of our two samples, new firms and diversification ventures with five or more employees, which allows us to break down the contributions to employment by types of new firms, including employee spinoffs. The contribution to 2001 employment of all the employee spinoffs that, from 1995 to 2001, entered with five or more employees and survived is 5.0 percent. This is 27.5 percent of the contribution of all entrants with five or more employees to Brazilian formal sector employment.[9]

4.6 Employee Spinoff Performance

Dunne et al. (1988) measure entrant performance by initial size, market share, and exit rates. We lack output or sales data needed to measure market share, but we can measure initial size using number of employees and compute average initial wage as an additional indicator of performance. Table 4.4 shows unconditional means for initial numbers of employees and average initial wages for the groups of entrants listed in Table 4.2, and cumulative exit rates after five years for the subsets of these groups consisting of entrants born in 1995 or 1996, following Dunne et al. (1988) who examine cumulative exit rates of entrants at five year intervals. We see that, for both the director/manager and five or more employees samples of entrants, unrelated new firms have the smallest initial size and highest exit rate. Diversification ventures or divestitures show the best performance for every indicator across both samples, and employee spinoffs are always intermediate except for having the lowest average initial wage in the director/manager sample.

In Table 4.5 we further investigate the relative performance of entrant groups by initial size and exit rate, the performance measures used by Dunne et al. (1988), by controlling for industry and cohort composition. Columns 1 and 2 cover entrants

[9] As we document in our working paper (Hirakawa et al., 2010), the sectoral distribution of new firms in Brazil is broadly consistent with worldwide survey evidence on entrepreneurship. Roughly half of new firms enter in commerce, repair services, hotels and restaurants. The next highest frequency of entry is observed in real estate activities and business services. The mix of employee spinoffs, divestitures and unrelated new firms is roughly similar across sectors.

that have at least one director or manager and columns 3 and 4 cover entrants with at least five employees. Size is measured by the log of the number of employees so in columns 1 and 3 we drop entrants with zero employees on December 31 of their birth years. The dependent variable in columns 2 and 4 is an indicator that takes the value of one for entrants born five years earlier that have exited and zero otherwise.[10] The key explanatory variables in these linear regressions are indicators for employee spinoff, divestiture, and diversification venture, alongside controls for 4-digit *CNAE* industry and cohort (entry year of firm or venture).[11] The omitted baseline entrant type is unrelated new firms. The exponential functions of the coefficients on the key indicator variables in columns 1 and 3 therefore show, within an industry and within a cohort, the ratios of the sizes of employee spinoffs, divestitures, and diversification ventures of existing firms to unrelated new firms. Similarly, the coefficients on the key indicator variables in columns 2 and 4 show, within an industry and within a cohort, the percentage point differences between the exit rates of unrelated new firms and those of employee spinoffs, divestitures, and diversification ventures of existing firms.

Table 4.5 shows that diversification ventures of existing firms are about three times as large as unrelated new firms among entrants with directors or managers and one and two-thirds as large among entrants with at least five employees. The first result accords especially well with the findings of Dunne et al. (1988) for U.S. manufacturing entrants, who state (p. 504) that "new-firm entrants in each industry are on average 28.4% as large as existing producers, while diversifying-firm, new-plant entrants are 87.1% … as large."[12] For both samples of entrants, divestitures are closer in size

[10]As explained at the end of Appendix 4.9.4, a new firm or venture is not considered to have exited until all its initial plants have exited. Even then, however, a new firm's 8-digit CNPJ root could survive because it has introduced a new plant. Survival of a firm's CNPJ root after exit of all its initial plants is very rare in our data. Modification of our exit definition for new firms to take account of this possibility causes the estimated exit rates for new ventures to rise relative to those for new firms by quantitatively insignificant amounts.

[11]The industry indicators used as controls in Tables 4.5 through 4.7 are based on the mode sector for new firms during their first year in the data.

[12]To make the comparison with Dunne et al. (1988) more accurate, we can drop the indicators for employee spinoff and divestiture from columns 1 and 3 of Table 4.5 so that the coefficients on diversification ventures give their sizes relative to all new firms, not just unrelated new firms. In this case the new coefficients for columns 1 and 3 are 2.66 and 1.58, respectively.

to diversification ventures than to unrelated firms, which supports our criteria for identifying divestitures since they should look like ventures of existing firms rather than new firms. Employee spinoffs, on the other hand, are much closer to the entry size of unrelated new firms than to diversification ventures of existing firms.

We also see from Table 4.5 that a diversification venture is 3 percentage points less likely to have exited than an unrelated new firm after five years for the director/manager sample and 8 percentage points less likely for the sample of entrants with at least five employees. The second result is consistent with the findings of Dunne et al. (1988, p. 513) for U.S. manufacturing entrants, who compute exit rates for diversification ventures from 6 to 14 percentage points lower than for new firms after five years, depending on cohort.[13] For both samples of entrants, divestitures have the lowest cumulative exit rate of any venture. Finally, manager and workforce employee spinoffs have cumulative exit rates after five years that are respectively 10 and 7 percentage points lower than those of unrelated firms.[14]

Our aim in this section is to establish regularities regarding the performance of employee spinoffs relative to other entrants, rather than test hypotheses about relative performance. Nevertheless, there is a mechanical reason why manager and especially workforce spinoffs should show better performance, and we would like to control for this. Application of both spinoff definitions requires that we be able to track employees at a new firm to previous employment. Mechanically, then, employees at a manager and especially workforce spinoff are more likely than employees at an unrelated new firm to have formal sector work experience. It would not be surprising if such firms were to survive in the formal sector longer. In the first and fifth columns of Table 4.6, therefore,

[13]As before, we can make the comparison with Dunne et al. (1988) more accurate by dropping the indicators for employee spinoff and divestiture from columns 2 and 4 of Table 4.5 so that the coefficients on diversification ventures give exit relative to all new firms, not just unrelated new firms. The new coefficients for columns 2 and 4 are then are then -.012 and -.055, respectively, the latter just missing the low end of the Dunne et al. (1988) range.

[14]A potential concern is that the superior performance of employee spinoffs relative to unrelated new firms is driven by firms with incorporated legal form, for which the classification of new firms as employee spinoffs is less certain. We reran our size and exit regressions for firms with non-incorporated legal form only, dropping divestitures. The differences in initial size and exit rates between employee spinoffs and unrelated new firms were qualitatively unchanged.

we add a control variable for the share of new entrant employees who are *trackable*, i.e., employees who had a formal sector job before. As expected, a greater share of trackable employees is associated with reduced cumulative exit rates for both entrants with at least one director or manager and entrants with at least five employees. However, the impact on exit rates of spinoffs is only slightly reduced.

Do larger initial sizes explain the lower cumulative exit rates of employee spinoffs (and divestitures and diversification ventures) relative to unrelated new firms? To answer this question we add the log of the number of initial employees as a control variable in columns 2 and 6 of Table 4.6.[15] This is indeed associated with lower exit rates for both entrants with at least one director or manager and entrants with at least five employees. The impacts on exit rates of divestitures and diversification ventures with at least five employees are slightly reduced, but the impact on the exit rate of employee spinoffs with at least five employees is unchanged. There are greater changes for entrants with at least one director or manager. For employee spinoffs and divestitures, impacts on exit rates are now below those for the same categories with at least five employees. For diversification ventures, the impact on exit rates is now slightly positive. Nevertheless, it is clear that the lower cumulative exit rates of employee spinoffs relative to unrelated new firms are an element of superior performance over and above greater entry size.

Given the findings for average initial wages in Table 4.4, it seems prudent to also control for human capital of startup employees when comparing cumulative exit rates of entrant types. An employee's wage at the preceding firm is a measure of the human capital that the employee brings to the current job. Indeed, the log average monthly wage that employees earned in their previous jobs shows a negative association with exit rates (statistically significant at the five-percent level for entrants with a director or manager), but other coefficient estimates are unaffected (columns 3 and 7). An employee's human capital includes his abilities as well as employer-related knowledge that he may transfer.

[15] Initial employees in these tables include all founding employees with a job at the new firm at any time during the first year, rather than in December only.

Related to this distinction, log wage components can be discerned in the annual cross sections of our linked employer-employee data: the employee component associated with individual observable characteristics on the one hand and an employer-specific component on the other hand, given a residual earnings component that includes the unobserved match effect.[16] In columns 4 and 8 we replace the employees' average log wages at their prior employer with the three log-wage components and observe that the individual worker characteristics component matters most. Employees with highly compensated observable characteristics at their previous employers are also significantly valuable in raising survival chances of entrants. In contrast, a high log-wage component of the previous employer is associated with higher exit rates (statistically significant at the five-percent level in the larger sample of entrants with five or more employees). A reason is perhaps that the presence of competing high-wage plants reduces an entrant's survival chance. Unobservable match characteristics at the previous employer are not significantly related to entrant performance.

Finally, it is possible that some of the apparently better performance of employee spinoffs relative to unrelated new firms results from an overly restrictive definition for divestitures. In other words, some employee spinoffs may actually be planned divestitures even though they contain less than 70 percent of the employees of any plant of their parent firm. To control for this possibility, we added a variable for the share of employees of the plant of the parent firm from which the entrant absorbs the most employees (not shown).[17] This variable has no statistically significant association

[16]Concretely, we decompose the log average monthly wage in a given year as

$$\ln w_i = x_i \beta + \psi_{J(i)} + \varepsilon_i$$

following Menezes-Filho et al. (2008), where w_i is employee i's annual wage, x_i is a vector of observable worker characteristics including gender, experience, education and occupation, β is a vector of estimated parameters, $\psi_{J(i)}$ is a plant effect ($j = J(i)$ being the plant that employs i), and ε_i is an error term. The plant effect combines a pure plant effect with the plant average of pure worker effects: $\psi_j = \phi_j + \overline{\alpha}_j$, where ϕ_j is the pure plant effect and $\overline{\alpha}_j$ is the average of pure employee effects α_i over employees at plant j (Abowd et al., 2001). Our decomposition thus attributes an employee's co-worker effects to the plant.

[17]For an unrelated new firm with at least one director or manager, a "parent" is just the existing firm from which the new firm received its top employee. For an unrelated new firm with at least five employees, a parent is just the existing firm from which the new firm absorbs the most employees, where "most" could

with cumulative exit rates, and the coefficients for employee spinoffs are essentially unaffected.

Exit does not necessarily imply failure. A new firm may be acquired by another firm and thereby earn its founders a tidy return. We define an exiting new firm or venture as *absorbed* if at least 70 percent of the exiting firm or venture's workforce is contracted by another firm during the year of exit; otherwise we call the exit a *failure*. For a meaningful application of the 70-percent definition, we restrict the sample to entrants with at least five employees at time of exit. When we restrict the regression sample to failures and survivors (dropping absorptions from the sample) in column 2 of Table 4.7, the inferior performance of unrelated new firms becomes even starker, and diversification ventures show the largest difference between failure rates and general exit rates. We restrict the regression sample to absorptions and survivors (dropping failures from the sample) in column 3. Compared to unrelated new firms, spinoffs and divestitures are more likely to be absorbed, and diversification ventures are three to four times more likely to be absorbed than are spinoffs or divestitures.[18]

We can summarize by returning to the interpretive framework we set out in our introduction. Of the four types of entrants, divestitures have the least product uncertainty, and we find that they have the lowest exit (or failure) rates. Diversification ventures of existing firms have the greatest access to the capital (human, organizational, and other) of the parent, and we find that they have the greatest size at entry. Employee spinoffs partially inherit, through embodiment in workers, the lower product uncertainty and various forms of capital of the parent, and are intermediate in exit rates between divestitures and unrelated new firms and intermediate in size at entry between diversification ventures (and divestitures) and unrelated new firms.

---

be as low as one.

[18]Of those entrants that are absorbed, 45 percent of diversification ventures are absorbed by their parents compared to 28 percent of spinoffs and 26 percent of divestitures.

4.7 Conclusion

Employee spinoffs have been found to be an important type of new business in many industries and many economies. Existing firms continuously lose employees, some of whom spin off to start their own businesses. Rich linked employer-employee data for Brazil allow us to systematically compare employee spinoffs to other new businesses, including management-initiated divestitures, and to diversification ventures of existing firms. Our identification of employee spinoffs draws on employer-reported occupations, firm identifiers and industry classifications, as well as firms' legal forms and mass employment shifts between firms.

Under one criterion, employee spinoffs are defined as new firms whose top salaried director or manager moved from a parent in the same industry. Under a second criterion, employee spinoffs are defined as new firms that fill at least a quarter of their jobs with employees who shifted from a common parent. Our findings are largely consistent across the two employee-spinoff definitions and lend mutual support to the definitions. Additional restrictions set employee spinoffs apart from divestitures and other entrants. Depending on definition, employee spinoffs account for between one-sixth and one-third of the respective new firms in Brazil's private sector during the period 1995-2001. Employee spinoffs grow into important employers. Total employment of employee spinoffs with at least a quarter of their workforce from a common parent, entering from 1995 to 2001, reaches five percent of all Brazil's formal-sector employment, private and public, by the end of the period.

Employee spinoffs are larger at entry than unrelated new firms but smaller than diversification ventures of existing firms. Similarly, employee spinoffs survive more frequently than unrelated new firms and with comparable frequency to diversification ventures. These results are consistent with the idea that employees embody some part of a parent firm's productivity and riskiness and that this capability is portable by the employees to a new firm. The literature on high-tech spinoffs has emphasized employee knowledge that is alienable intellectual property. We found that the bulk of employee spinoffs is in non-high tech sectors, and that on average top managers bring 23 percent of the spinoff workforce with them from parent firms. Both facts suggest that knowledge that is tacit or at least not easily contractible is an important factor in the success of most employee spinoffs.

Our findings have potentially important implications beyond firm dynamics and entrepreneurial policy. For example, using our quarter-workforce spinoff definition, Muendler and Rauch (2011) identify parent firms that spawn both employee spinoff plants and expansion or diversification plants, and show that spinoff plants locate even closer to their parents than the parents' own new plants, controlling for sector, the share of initial employees from the parent, and initial plant size. This supports the argument of Klepper (2010), based on case studies of the U.S. automobile industry in Detroit and the U.S. semiconductor industry in Silicon Valley, that employee spinoffs can play a key role in the initiation of industry clusters. We hope that our quantification of the employee spinoff phenomenon across multiple industries encourages further research into all its impacts.

Acknowledgements

We have benefited from comments by Victor Gomes, John Haltiwanger, Daniel Lederman, the editor and referees, and seminar participants at Georgia Tech, New York University, Universidade de Brasília, UC Davis, and the World Bank. We acknowledge support from the International Growth Centre (RA-2010-03-002) and the World Bank with gratitude. Muendler also acknowledges NSF support (SES-0550699) with gratitude.

4.8 Tables

Table 4.1: Samples of entrants for analysis

Type of new firm or venture	Count	Employment (thousands)
Universe of new firms	1,515,560	6,038
with director/manager	5.0%	25.5%
with five or more employees	21.9%	73.0%
Universe of diversification ventures	60,593	1,365
with director/manager	10.9%	54.5%
with five or more employees	48.4%	96.4%

Source: RAIS 1995-2001, new firms and diversification ventures of existing firms.
Notes: A new firm is a firm in 1995-2001 whose tax ID did not exist in RAIS 1986-1994 (see Appendix 4.9.2 for tax ID assignment). A diversification venture consists of all new plants started by an existing firm if these plants are in a different *CNAE* 4-digit industry from the existing firm, comparing the industry associated with the top employee at a new plant with the mode sector of the existing firm in the previous year. The universe does not include new firms of excluded legal form (28,087 firms with 1,041 thousand employees), diversification ventures of excluded legal form (1,751 ventures with 190 thousand employees) and undetermined sector ventures (20,036 ventures with 111 thousand employees). An undetermined sector venture collects an existing firm's new plants whose industry cannot be compared to the original firm's industry because of missing information for new plant or original firm. Employment measured in December.

Table 4.2: Entrants by sample and type

Type of new firm or venture	Count	Employment (thousands)
Director/Manager Sample		
New firms, *of which:*	76,497	1,542
Employee spinoffs	17.0%	23.9%
Divestitures	4.3%	22.2%
Unrelated new firms	78.7%	53.9%
Diversification ventures	6,582	744
Five or More Employees Sample		
New firms, *of which:*	331,987	4,409
Employee spinoffs	29.3%	31.9%
Divestitures	5.5%	14.9%
Unrelated new firms	65.3%	53.1%
Diversification ventures	29,348	1,315

Source: RAIS 1995-2001, new firms and diversification ventures of existing firms.
Notes: Types of new firms defined as in Table 4.9, Appendix 4.9.4. Diversification ventures defined as in Table 4.1. Employment measured in December.

Table 4.3: Shares of formal sector employment by entrant, 1995-2001

	1995	1998		2001	
	current	current	cumulative[a]	current	cumulative[a]
RAIS universe	23,222	24,606		27,426	
of which:					
New firms	2.8%	3.5%	13.0%	3.4%	22.2%
Diversification ventures	0.9%	0.8%	2.5%	0.7%	3.4%
New firms (5+ employees)	2.0%	2.6%	8.9%	2.5%	14.9%
Spinoffs	0.6%	0.9%	2.9%	0.8%	5.0%
Divestitures	0.3%	0.4%	1.1%	0.4%	2.0%
Unrelated	1.1%	1.3%	4.9%	1.3%	8.0%
Div. ventures (5+ employees)	0.9%	0.8%	2.3%	0.7%	3.3%

[a] Includes the 1998 (2001) employment of new firms and ventures born between 1995 and 1998 (2001). Only the entrants' original plants are included, so the cumulative shares underestimate slightly the importance of new entrants.
Source: RAIS 1995-2001 universe of all formal sector firms, including otherwise removed new firms with inadmissible legal form.
Notes: Types of new firms defined as in Table 4.9, Appendix 4.9.4. A diversification venture is defined as in Table 4.1. Employment measured in December in thousands.

Table 4.4: Summary statistics

Type of New Firm or Venture	Initial employment	Average initial wage (BRL)	Exit after 5 years
Director/Manager Sample			
New firms	21.25	372.00	52.9%
	(0.63)	(2.10)	(0.4%)
Employee spinoffs	29.32	344.20	45.0%
	(1.07)	(4.28)	(1.0%)
Divestitures	106.93	406.36	39.7%
	(8.87)	(10.00)	(2.1%)
Unrelated new firms	14.62	376.22	55.1%
	(0.58)	(2.45)	(0.5%)
Diversification ventures	117.15	491.28	46.7%
	(6.72)	(7.45)	(1.2%)
Five or More Employees Sample			
New firms	13.86	221.09	45.9%
	(0.16)	(0.42)	(0.2%)
Employee spinoffs	15.01	254.29	39.0%
	(0.29)	(0.95)	(0.3%)
Divestitures	37.37	250.39	38.1%
	(2.22)	(2.10)	(0.8%)
Unrelated new firms	11.33	203.53	49.4%
	(0.09)	(0.45)	(0.2%)
Diversification ventures	45.76	257.08	28.5%
	(1.54)	(1.94)	(0.5%)

Source: RAIS 1995-2001, new firms and diversification ventures of existing firms with at least one director/manager or at least five employees.

Notes: Types of new firms defined as in Table 4.9, Appendix 4.9.4. Diversification ventures defined as in Table 4.1. Employment measured in December. December wages in BRL, deflated to the August 1994 price level. Standard errors in parentheses.

Table 4.5: Size at entry and cumulative exit five years after entry

	Director/manager		Five or more employees	
OLS exponentials of coefficients in columns 1 and 3	log Empl. at t (1)	Exit by $t+5$ (2)	log Empl. at t (3)	Exit by $t+5$ (4)
Employee spinoff	1.85	-.096	1.12	-.069
	(.022)**	(.012)**	(.004)**	(.004)**
Divestiture	2.67	-.151	1.41	-.121
	(.074)**	(.022)**	(.011)**	(.008)**
Diversification venture	3.11	-.033	1.67	-.082
	(.066)**	(.014)*	(.014)**	(.007)**
Obs.	78,911	16,564	346,813	87,476
R^2	.29	.08	.13	.10
Mean Dep. variable	1.75	.52	2.07	.44
CNAE industry panels	550	504	560	538
Cohort panels	7	2	7	2

Source: RAIS 1995-2001, new firms and diversification ventures of existing firms with at least one director/manager or at least five employees.
Notes: Types of new firms defined as in Table 4.9, Appendix 4.9.4. Diversification ventures defined as in Table 4.1. Omitted category: unrelated new firms. Employment measured in December. Coefficients in columns 1 and 3 reported as exponential functions of coefficients from OLS regression so they reflect the ratio of sizes relative to unrelated new firms; standard errors in columns 1 and 3 computed with the Delta method. All regressions condition on *CNAE* industry and cohort fixed effects. Robust standard errors in parentheses: * significance at five, ** at one percent level.

Table 4.6: Cumulative exit five years after entry: additional specifications

	Director/manager				Five or more employees			
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Employee spinoff	-.086	-.060	-.064	-.062	-.066	-.069	-.069	-.063
	(.012)**	(.012)**	(.012)**	(.012)**	(.004)**	(.004)**	(.004)**	(.004)**
Divestiture	-.137	-.096	-.103	-.096	-.119	-.116	-.116	-.110
	(.022)**	(.022)**	(.022)**	(.022)**	(.008)**	(.008)**	(.008)**	(.008)**
Diversification venture	-.025	.031	.026	.027	-.081	-.070	-.069	-.063
	(.014)	(.014)*	(.014)	(.014)	(.007)**	(.007)**	(.007)**	(.007)**
Share: Trackable	-.077	-.051	-.025	-.022	-.014	-.002	-.005	-.0003
	(.014)**	(.014)**	(.021)	(.021)	(.008)	(.008)	(.008)	(.009)
log Initial empl.		-.050	-.048	-.053		-.024	-.024	-.025
		(.003)**	(.003)**	(.003)**		(.002)**	(.002)**	(.002)**
Prev. log Wage			-.020				-.007	
			(.006)**				(.004)	
Indiv. component				-.088				-.117
				(.017)**				(.010)**
Plant component				.012				.019
				(.010)				(.005)**
Residual				-.023				-.002
				(.013)				(.008)
Obs.	16,564	16,564	15,224	15,224	87,476	87,476	85,894	85,894
R^2	.083	.098	.098	.099	.098	.099	.100	.101
Mean dep. var.	.52	.52	.51	.51	.44	.44	.44	.44
CNAE ind. panels	504	504	502	502	538	538	538	538
Cohort panels	2	2	2	2	2	2	2	2

Source: RAIS 1995-2001, new firms and diversification ventures of existing firms with at least one director/manager or at least five employees.
Notes: Types of new firms defined as in Table 4.9, Appendix 4.9.4. Diversification ventures defined as in Table 4.1. Omitted category: unrelated new firms. The previous log wage is the average monthly log wage (BRL deflated to August 1994) at the preceding employer; the three log wage components—individual characteristics, the plant-fixed component, and a residual—are from a linear regression decomposition of the previous log wage into employee observables, a plant fixed effect and a residual using repeated annual cross sections of the linked employer-employee data. All regressions condition on *CNAE* industry and cohort fixed effects. Robust standard errors in parentheses: * significance at five, ** at one percent level.

Table 4.7: Cumulative failure and absorption five years after entry: five or more employees sample

	Any Exit	Failure	Absorption
OLS	(1)	(2)	(3)
Employee spinoff	-.069	-.082	.024
	(.004)**	(.004)**	(.002)**
Divestiture	-.121	-.138	.024
	(.008)**	(.008)**	(.005)**
Diversification venture	-.082	-.120	.081
	(.007)**	(.007)**	(.006)**
Obs.	87,476	84,784	51,686
R^2	.098	.104	.043
Mean Dep. Variable	.44	.42	.05
CNAE industry panels	538	537	528
Cohort panels	2	2	2

Source: RAIS 1995-2001, new firms and diversification ventures of existing firms with at least five employees.
Notes: Column 1 restates results from Table 4.5 (column 4), subsample in column 2 excludes absorptions, subsample in column 3 excludes failures. An absorption is defined as the shift of 70 percent of an exiting firm's workforce to another firm (firm exit is failure otherwise). A new firm is defined as in Table 4.9, Appendix 4.9.4. A diversification venture is defined as in Table 4.1. Omitted category: unrelated new firms. All regressions condition on *CNAE* industry and cohort fixed effects. Robust standard errors in parentheses: * significance at five, ** at one percent level.

4.9 Appendix

4.9.1 Employer-employee Data

Screening of employee data. Employees in RAIS are identified by the individual-specific PIS number (*Programa de Integração Social)*. A given plant may report the same PIS multiple times within a single year so that the employee can withdraw from the employer-funded severance pay account (*FGTS*) through spurious layoffs and rehires. In addition, some PIS values (especially very small or symmetric numbers) are recorded by an unrealistically large number of different plants. To handle these issues, we devise a systematic way to label PIS values that we think should not be trusted for tracking employee's employment histories: if an employee appears at more than twelve jobs in any given year, or if there is more than one apparent gender change (i.e. there are two or more years in the data when the employee is listed as being of both genders), we mark the employee has having an invalid PIS. None of the 14,272 employees caught by this rule is deleted from the data. Instead, we only disregard their work history for purposes of identifying the parent of a new firm and for defining spinoffs.

To avoid double-counting employees at new firms, we keep only one observation for each employer-employee-year combination, choosing the job with the earliest hiring date. If the employee has two jobs at the firm starting in the same month, we keep the highest paying one (randomly dropping all but one observations with equal monthly average wages). For new ventures of existing firms, we apply this rule at the plant-year level, thus allowing the employee to appear once per plant during the plant's first year, again choosing the job with the earliest hiring date and highest average monthly wage.

To compute the December performance measures (employment and wage per worker) as reported in Tables 4.1 and 4.2, and as employed on the left-hand side in Table 4.5, we choose a modified version of the data cleaning described above. Instead of allowing only one observation per employee per year at the new firm or plant, we allow only one observation per employee on December 31 at the given firm or plant (in the job with the top December wage). This way we make sure that we do not lose from our December count any employees who worked in a different occupation at the firm earlier in the year.

Earnings. An employee's earnings in RAIS are expressed in multiples of the monthly minimum wage that prevails at the time. RAIS reports the average monthly wage during a calendar for all observations, and the December wage for jobs held at year end. We use both wage measures, depending on context, and calculate the wage value in Brazilian Real (BRL), deflated to August 1994. In July 1994, Brazil adopted a new monetary regime with single-digit annual inflation rates (starting with a BRL value at par with the U.S. dollar).

The RAIS manual for respondents states explicitly the forms of payment that are considered valid components of the monthly wage rate. These include: salaries; extraordinary additions, supplements and bonuses; tips and gratuities; commissions and fees; contracted premia; overtime earnings for contracted extra hours; hazard earnings; executive earnings; cost reimbursement components if they exceed fifty percent of the base salary and are for travel or transfers necessary for the execution of the job; payments for periods of vacation, holidays and parental leave; vacation gratuities if they exceed twenty days of salary; piece wages; and in-kind remunerations such as room and board. As a rule, components are considered part of salary if they are taxable income or are subject to Brazilian social security contributions.

Payments that are not considered wage components include: severance payments for layoffs; indemnity payments for permanent maternal leave and any other indemnity payments; so-called "family payments" under Brazilian labor law; vacation gratuities if they do not exceed twenty days of salary; additional social security earnings due to an employee's illness; moving expenses; travel cost reimbursements if they do not exceed fifty percent of the base salary; scholarships for interns; meals, equipment and clothing for execution of the job; participation in the employer's profits; and so-called pro-labore payments for services by owners who do not have a dependent employment relationship.

Occupations. Occupations are categorized using the so-called *CBO* classification codes in RAIS. For our implementation, it is not necessary to reclassify *CBO* codes to conform with the *ISCO-88* categories. Our main use of the occupational coding is to identify directors and managers. The Portuguese title 'diretor geral', for instance, is similar to the occupation of a CEO, 'diretor de financas' similar to CFO.

4.9.2 Firm Identifiers

Consistent application of firm identifiers is crucial for our identification of new plants and firms. Plant-level information in RAIS is based on the CNPJ identification number, where CNPJ ('cadastro nacional de pessoa juridica') stands for Brazil's national register of legal juristic persons. The first eight digits of CNPJ numbers (CNPJ radical) define the firm and the subsequent six digits the plant/branch within the firm. The CNPJ number is assigned or extinguished, and pertaining register information updated, under legally precisely defined conditions.

The CNPJ number is administered by the Brazilian tax authority Receita Federal, the Brazilian equivalent to the U.S. IRS. In the CNPJ register, Receita Federal maintains information related to the firm's legal form and related matters, which is separately also recorded in RAIS. The following nine types of transactions either trigger the creation or extinction of CNPJ numbers, or updating of the register while maintaining CNPJ numbers. Once extinguished, a CNPJ number cannot be reassigned to any other plant in the future.

1. *Opening a business, becoming a juristic person.* Obtain CNPJ. It is required of any juristic person ('pessoa juridica') in Brazil, a legal entity in Brazilian common and commercial law, to register a CNPJ number with the Receita Federal upon opening a business.[19]

[19]There is also a set of legal entities that are not formally juristic persons but are put on equal legal

2. *Change in business name ('nome empresarial'), or business sector ('porte da empresa'), or legal form ('natureza juridica').* Maintain CNPJ, update register information. Changes from individual entrepreneurs to associations or partnerships of entrepreneurs and owners, or the reverse, do not result in reported changes in legal form.

3. *Change in ownership ('quadro de sócios') at associations and partnerships, or change in management ('administradores'), or change in equity holding at associations and partnerships ('inclusão e alteração de capital social').* Maintain CNPJ, update register information. Note that changes to incorporated firms—juristic persons with independent legal existence such as a limited liability company ('sociedade por quotas de responsabilidade limitada')—are treated differently, see 8 below.

4. *Other changes to the register, including mothballing ('interrupção temporária de atividades') and resumption of operations ('reinício das atividades interrompidas temporariamente'), a change in tax status ('opção ou exclusão do simples', 'qualificação tributária'), a change of responsible physical person (human being) for the CNPJ juristic person ('pessoa física responsável perante o CNPJ'), and several other administrative cases.* Maintain CNPJ, update register information.

5. *Bankruptcy and liquidation.* Maintain CNPJ, update register information. It pertains to the Receita Federal to administer the CNPJ of the extinguished juristic person. Liquidation may be by court order or extrajudicial settlement. The opening and closing of a bankruptcy case must be reported.

6. *Opening new plants/branches.* New plants or branches are registered with the individual CNPJ numbers, where the first eight digits (CNPJ radical) define the firm and the subsequent six digits the plant/branch within the firm.

footing with juristic persons by Receita Federal, including real estate condominiums, mutual funds, employer consortia, and foreign consulates.

7. *Partial divestiture/corporate spinout ('cisão parcial').* Maintain CNPJ, update register information. The newly independent firm (divestiture or spinout) receives an own CNPJ. In practice, a partial divestiture might coincide with the acquisition of an individual plant by another firm.

8. *Merger of firm with other firm ('fusão'), acquisition of firm by other firm ('incorporação') or complete divestiture/corporate spinout into newly independent firms ('cisão total').* Extinguish CNPJ of firm that undergoes change. In the case of mergers and complete divestitures, the newly independent firm(s) obtain CNPJ(s) of their own. In the case of a plant acquisition, if the divested plant is not incorporated as a firm, the acquiring firm's CNPJ radical is retained and six new digits for the new plant are added. Note that the above applies to the acquisition of the firm as a whole, not select plants within the firm (for those cases see 7).

9. *Inactivity since day of foundation ('empresa que não iniciou atividades (inativa desde a abertura)').* Extinguish CNPJ.

Important for employee spinoffs, a change in ownership at associations or partnerships does not result in a change in CNPJ, as explained under item 3. Divestitures include both management-initiated offspring that become standalone firms (corporate spinouts or complete splitups ('cisão total')) and management-initiated offspring from parent firms' M&A activity (such as a merger ('fusão'), an acquisition ('incorporação'), and a partial splitup ('cisão parcial')). These are covered under items 7 and 8.

4.9.3 Natureza Juridica (Legal Form)

Employee spinoffs are employee-initiated offspring firms whose key employees stem from one or multiple legally separate parent firms. We choose our empirical implementation such that it is unlikely that parent firms or acquiring companies hold a capital stake in the employee spinoff (the employee spinoff may or may not face contractual obligations with the parent firm). For this purpose, we use the *natureza juridica* (legal form) variable in RAIS to discern three important types of legal form: associations or partnerships without independent legal existence, private incorporated firms, and types of incorporated firms to be excluded from analysis. Associations or partnerships can only be owned by physical persons, not by other companies. There is minor reporting error in legal form: around .1 percent of new firms have more than one (non-missing) legal form in their first year. We assign the mode of its legal form during the year to every firm.

Table 4.8: Treatment of legal form

Natureza Juridica (legal form)	Presumed type: Non-incrp.	Presumed type: In-corp.	Presumed type: Ex-cluded	*Total*		RAIS codes
Public administration			x	6,718	.4%	1015-1996
State-owned company[a]			x	16,909	1.1%	2011-2038
Corporation		x		4,110	.3%	2046, 2054
Limited liability company		x		867,656	56.2%	2062
Partnership	x			3,008	.2%	2070-2100, 2127
For-profit association	x			47,193	3.1%	2119
Sole-proprietor company[b]	x			493,130	32.0%	2135, 2992
Cooperative			x	3,553	.2%	2143
Consortium			x	318	.02%	2151
Business group			x	436	.03%	2160
Branch of foreign company			x	153	.01%	2178
Non-profit organization	x			77,616	5.0%	3018-3999
Professional w/out employees[c]	x			379	.02%	4030
Professional w/ employees[c]	x			4,880	.3%	4049
Entrepreneurial proprietor	x			1,518	.1%	4073
Other professional[c]	x			2,408	.2%	4014-4995[d]
Unknown		x		13,662	.9%	.
Total				1,543,647	100.0%	

[a] State-owned limited liability company and close corporation, and Corporation with some state control.
[b] Includes other private businesses.
[c] Includes self employment.
[d] Excluding above codes.
Source: RAIS 1995-2001, new firms.
Note: Incorporated legal forms underly the definition of a divestiture (a new firm with *natureza juridica* coded as Corporation under private control, Close corporation, Limited liability company, or as unknown that absorbs 70 percent or more of the employees of a plant of an existing firm). Excluded legal forms are Branches of government, Firms with state ownership, Cooperatives, any type of Holding company, and Branches of foreign firms.

Table 4.8 shows the frequency of *natureza juridica* among new firms. More than 97 percent of new firms are concentrated in just four legal forms: limited liability companies with 56 percent, sole-proprietor companies with 32 percent, non-profit organizations (5 percent) and for-profit associations (4 percent). Only the limited liability company is an incorporated legal type that can be owned by another company, whereas the remaining three legal forms among the top four are associations or partnerships without independent legal existence. As mentioned, associations or partnerships can only be owned by physical persons. The latter three legal forms are thus also not subject to CNPJ changes, see item 3 in the preceding Appendix. We consider the latter three legal forms highly likely employee spinoffs if they satisfy the criteria of our manager or workforce definitions. We return to the use of *natureza juridica* in our description of spinoff and divestiture definitions below.

4.9.4 Implementation of Spinoff and Divestiture Definitions

We apply two distinct sets of spinoff criteria (our manager and workforce definitions), each administered at the firm level (first eight digits of the CNPJ tax number). To identify a potential parent firm, we use the job histories of the new firm's founding employees, where the *founding employees* are the individuals employed at the firm during its first year in RAIS.[20] In particular, for each of the founding employees we identify the previous *substantial job* as the last preceding employment spell (by hiring month) with a duration of at least three months.[21] We search for the previous job as far back as the RAIS data allow us. Our data start in 1986, which gives us nine years of potential labor market experience before 1995, the year in which we first consider firm entries.

[20]Firm age comparisons with other data sources show that RAIS reports date of firm creation plausibly precisely.

[21]If the employee started two or more jobs in a month, we select the highest paying job, randomly dropping ties. We also require that the previous employment spell is at a different firm than the new firm at which the employee is currently employed.

Manager spinoff. The *director/manager* criterion isolates the top employee at each new firm first by job description (where director trumps manager, which trumps other descriptions), and secondarily by average monthly wage. The previous firm at which this top employee worked for at least three months is identified as the new firm's parent. If this parent is within the same disaggregated industry (same 4-digit *CNAE* industry of which there are 654) as the new firm, and the top employee is a director or manager, we label the new firm a spinoff. For this purpose, we do not compare mode industries of the parent and new firm (since the parent firm may operate plants in several industries); instead we use the industries associated with the transferring top employee at her old and new job. If either of the two industries is missing, the spinoff definition is not satisfied. If there are two or more manager employees tied for top employee, the firm is labelled a spinoff if any one (or all) of these employee's parent firms is in the same industry as the new firm. So multi-parent spinoffs are possible, but they are rare in practice (multi-parent spinoffs represent 0.7 percent of all manager spinoffs). This definition is only applied to new firms with management-level employees, about 5 percent of the entire new-firm sample (see Table 4.1).

Workforce spinoff. The *workforce* definition considers the previous place of substantive employment (lasting at least three months) of all the new firm's employees, regardless of job description or pay. The parent firm is the firm that supplied the largest number of employees to the new firm. The new firm is labelled a spinoff as long as 25 percent or more of the new firm's employees come from the parent firm. This definition would trivially label as spinoffs all firms with four or fewer initial employees, therefore we only apply it to the new firms with five or more initial employees. Multi-parent spinoffs are again possible (they constitute 4.7 percent of the workforce spinoffs).

For both spinoff definitions, if there are two or more parent firms (multi-parent spinoff), we keep the parent within the same industry for purposes of testing the mass employee shift criterion for divestitures. Any remaining ties are broken at random to select a unique parent.

Legal form of new firm. We further use legal form data (the mode calculated for each new firm) to help distinguish employee spinoffs from management-initiated divestitures. As described above (Appendix 4.9.3), incorporated firms can be owned by other companies and can thus be subject to CNPJ changes as ownership changes (Appendix 4.9.2). For new firms that are incorporated, management-initiated divestitures could therefore be a motive of their creation (natureza juridica 2046, 2054 or 2062, or unknown). In contrast, personal businesses such as associations and partnerships cannot be owned by other companies under Brazilian commercial law, and are thus not subject to CNPJ changes. We therefore consider associations and partnerships as highly likely employee spinoffs if they satisfy the spinoff definitions (natureza juridica 2070-2135, 2992, 3018-3999, 4014-4995). We exclude from the analysis legal forms that designate employers as public administration (natureza juridica 1015-1996), state-owned companies or corporations with some state control (2011-2038) or as special companies such as cooperatives, consortia, business groups and branches of foreign companies (2143-2178). Table 4.8 documents that the bulk of new firms' legal forms are included: 56.5 percent of new firms fall under the incorporated legal forms and 40.8 percent of new firms fall under the non-incorporated legal forms.

We apply the following refinement to our two spinoff definitions. A firm is a spinoff if a spinoff definition is satisfied (manager or workforce) and the legal form of the new firm is non-incorporated. A firm is also a spinoff if a spinoff definition is satisfied, the legal form of the new firm is incorporated, and strictly less than 70 percent of any parent plant's workforce shifts to the new firm. We now turn to the latter mass-employee shift criterion that distinguishes spinoffs from divestitures.

Divestitures, including corporate spinouts. If 70 percent or more of a parent plant's workforce switch to a new CNPJ from one year to the next, we call the new plant a divestiture plant. We impose no minimum size on a parent firm for this computation. This definition is based on an employee count at the parent, contrary to our spinoff definitions which are based on employee counts at the new firm. In particular, we identify the parent at the firm level and single out the parent-firm's plant with the highest fraction of employees that shift to a new firm. The denominator in the share of shifting employees is the count of substantive parent employees over the year prior to the new firm's entry.[22] If the new firm has no trackable employees, or if the parent firm did not appear in RAIS during the previous year, we cannot calculate the share of parent plant employees that shifted, and we assume that the value is below 70 percent.

The 70-percent cutoff is motivated by the reverse of the labor economists' definition of a mass layoff (e.g. Jacobson et al., 1993), by which 30 percent or more of the existing workforce experience a separation. We label all divestiture firms that originate from 70 percent of a parent plant's workforce with an according indicator in the data. So, we call a firm a divestiture if the legal form of the new firm is incorporated and at least 70 percent of the parent plant's workforce switch to the new firm. The share of parent plant employees that shift to the new firm is also used as an added control in exit probability regressions. For those regressions, we also need to construct the share of shifting employees at new ventures of existing firms. For new ventures, the parent firm is simply the 8-digit root part of the existing firm's CNPJ number. Similar to divestitures, we select the parent plant with the highest share of its employees lost to the new venture to calculate the denominator for the share of shifting employees.

Unrelated startup firms. Firms with included legal form that do not fall into the spinoff or divestiture categories are in the outside comparison group.

Table 4.9 in summarizes the exhaustive and mutually exclusive classification of new firms resulting from these definitions.

New ventures of existing firms. During our sample period 1995-2001, 580,557 new plants are started at 152,694 existing firms. We divide these into expansion plants (same 4-digit *CNAE* industry as parent firm), diversification plants (different 4-digit *CNAE*

[22]We count parent plant employees as follows. We disregard employment spells of less than three months, and we keep only one appearance of any given employee per year per plant.

Table 4.9: Classification of new firms

Type of New Firm	Spinoff criteria[a]	Mass Employee Shift[b]	Legal Form of New Firm[c]
Unrelated new firm	no	yes or no	non-incorporated
Unrelated new firm	no	no	incorporated
Employee spinoff	yes	yes or no	non-incorporated
Employee spinoff	yes	no	incorporated
Divestiture	yes or no	yes	incorporated
Excluded legal form	yes or no	yes or no	—

[a] See Appendix 4.9.4 for definitions of manager spinoffs and workforce spinoffs, based on a director/manager criterion (top employee switches from same-industry parent to spinoff) and on a quarter-workforce criterion (25 percent or more of the new firm's employees from same parent firm).
[b] See Appendix 4.9.4 for the criterion of a shift (70 percent or more of a parent plant's workforce switch) and the definition of a divestiture.
[c] For our classification into non-incorporated legal forms, incorporated legal forms and excluded legal forms, see Table 4.8 in Appendix 4.9.3.
Note: New firms are firms in 1995-2001 whose root tax ID (first eight digits) did not exist in RAIS 1986-1994. Legal form according to *natureza juridica* in RAIS.

industry), and plants for which we cannot perform the sector comparison (because either the new plant or the parent firm has no known sector). The parent firm's industry is the firm's mode *CNAE* sector during the immediately preceding year in the data. A diversification venture of an existing firm is the sum of its diversification plants. Analogously to new firms, a new venture passes the director/manager filter if any of its plants has a director or manager, and a new venture passes the five or more employees filter if the sum of its plants has five or more employees.

Mode sector assignment. For regression purposes, we assign to each firm (or plant) its mode sector value for that year, computed over the raw data and over all employees (not just December-31 employees). Many firms with no employees in December of a given year go on to have a workforce in December of future years. Of the new firms from 1995 that survive through 2001, for instance, more than seven percent had zero employment on December 31 of 1995. We would lose many observations in performance regressions controlling for initial year sector if we only based the sector on December-31 employees. For new ventures of existing firms, we compute the mode sector as follows: we take the mode sectors of its plants, weight them by the number of employees of each plant, and compute the mode. New firms or ventures with no known sector are not excluded from regressions, instead they are included under a common "unknown sector" category.

Exit. We adopt the following exit definition for the regressions in Tables 4.5 through 4.6: a plant is considered *active* (has not yet exited) in a year t if it has any employment at any time during year t or during any of the following years $t+\tau$. A new firm or venture survives as long as any of its initial plants is still active. We define the exit indicator variable $exit(t+\tau)$ to be 0 if the new firm or venture has not yet exited at year $t+\tau$, and to be 1 if it exited in $t+\tau$ or in a previous year. The exit indicator is only defined for firms and ventures for which it is possible to test survival. In particular, since our data end in 2001, $exit(t+5)$ is only defined for firms and ventures that enter in 1995 or 1996.

4.10 Bibliography

Abowd, John M., Francis Kramarz, David N. Margolis, and Kenneth R. Troske, "The Relative Importance of Employer and Employee Effects on Compensation: A Comparison of France and the United States," *Journal of the Japanese and International Economies*, December 2001, *15* (4), 419–436.

Alchian, Armen A. and Harold Demsetz, "Production, Information Costs, and Economic Organization," *American Economic Review*, December 1972, *62* (5), 777–795.

Anton, James J. and Dennis A. Yao, "Start-ups, Spin-offs, and Internal Projects," *Journal of Law, Economics, & Organization*, October 1995, *11* (2), 362–378.

Benedetto, Gary, John Haltiwanger, Julia Lane, and Kevin McKinney, "Using Worker Flows to Measure Firm Dynamics," *Journal of Business and Economic Statistics*, July 2007, *25* (3), 299–313.

Cabral, Luís M. B. and Zhu Wang, "Spin-offs: Theory and Evidence," *Federal Reserve Bank of Kansas City Research Working Paper*, December 2008, *RWP 08-15.*

Cusatis, Patrick J., James A. Miles, and J. Randall Woolridge, "Restructuring through Spinoffs: The Stock Market Evidence," *Journal of Financial Economics*, June 1993, *33* (3), 293–311.

De Negri, João Alberto, Paulo Furtado, Natália Ribeiro de Souza, and Jorge Saba Arbache, "Mercado Formal de Trabalho: Comparação entre os microdados da RAIS e da PNAD," *IPEA Texto para Discussão*, November 1998, *840.* Instituto de Pesquisa Econômica Aplicada, Rio de Janeiro.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson, "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries," *RAND Journal of Economics*, Winter 1988, *19* (4), 495–515.

Eriksson, Tor and Johan Moritz Kuhn, "Firm Spin-Offs in Denmark 1981-2000: Patterns of Entry and Exit," *International Journal of Industrial Organization*, September 2006, *24* (5), 1021–1040.

Franco, April Mitchell and Darren Filson, "Spin-Outs: Knowledge Diffusion through Employee Mobility," *RAND Journal of Economics*, Winter 2006, *37* (4), 841–860.

Hannan, Michael T. and John Freeman, "Structural Inertia and Organizational Change," *American Sociological Review*, April 1984, *49* (2), 149–164.

Henderson, Rebecca M. and Kim B. Clark, "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms," *Administrative Science Quarterly*, March 1990, *35* (1), 9–30.

Hirakawa, Oana, Marc-Andreas Muendler, and James E. Rauch, "Employee Spinoffs and Other Entrants: Stylized Facts from Brazil," *IGC Working Paper*, December 2010, *10/0879.*

Holmstrom, Bengt, "Moral Hazard in Teams," *Bell Journal of Economics*, 1982, *13* (2), 324–340.

Hvide, Hans K., "Firm Size and the Quality of Entreprenuers," *CEPR Discussion Paper*, 2005, *4979.*

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan, "Earnings Losses of Displaced Workers," *American Economic Review*, September 1993, *83* (4), 685–709.

Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica*, May 1982, *50* (3), 649–670.

Klepper, Steven, "Employee Startups in High-Tech Industries," *Industrial and Corporate Change*, September 2001, *10* (3), 639–74.

_ , "The Origin and Growth of Industry Clusters: The Making of Silicon Valley and Detroit," *Journal of Urban Economics*, January 2010, *67* (1), 15–32.

_ **and Sally Sleeper**, "Entry by Spinoffs," *Management Science*, August 2005, *51* (8), 1291–1306.

Menezes-Filho, Naércio Aquino, Marc-Andreas Muendler, and Garey Ramey, "The Structure of Worker Compensation in Brazil, with a Comparison to France and the United States," *Review of Economics and Statistics*, May 2008, *90* (2), 324–346.

Minniti, Maria, William D. Bygrave, and Erkko Autio, *Global Enterpreneurship Monitor 2005 Executive Report*, London: Global Entrepreneurship Research Association, 2005.

Muendler, Marc-Andreas and James E. Rauch, "Employee Spinoffs, Market Cannibalization, and Agglomeration: Evidence from Brazil," January 2011. University of California, San Diego, unpublished manuscript.

Nandy, Debarshi and Thomas Chemmanur, "How is Value Created in Spin-Offs? A Look Inside the Black Box," *U.S. Census Bureau Center for Economic Studies Working Paper*, July 2005, *05-09.*

Reynolds, Paul D., Michael Hay, William D. Bygrave, S. Michael Camp, and Erkko Autio, *Global Enterpreneurship Monitor 2000 Executive Report*, London: Global Entrepreneurship Research Association, 2000.

Shieh, Gwo-shyong, *"Boss" island: The subcontracting network and microentrepreneurship in Taiwan's development*, New York: Peter Lang, 1992.

Teece, David J., "Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy," *Ricerche Economiche*, October-December 1986, *40* (4), 607–43.

Wiggins, Steven N., "Entrepreneurial Enterprises, Endogenous Ownership, and the Limits to Firm Size," *Economic Inquiry*, January 1995, *33* (1), 54–69.