Essays on the Macroeconomics of Economic Development

by

Michael Peters

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2012

Author ....
Department of Economics
May 15, 2012

Certified by ....
Daron Acemoglu
Charles P. Kindleberger Professor of Applied Economics
Thesis Supervisor

Certified by ....
Abhijit Banerjee
Ford International Professor of Economics
Thesis Supervisor

Certified by ....
Robert M. Townsend
Elizabeth & James Killian Professor of Economics
Thesis Supervisor

Accepted by ....
Michael Greenstone
3M Professor of Environmental Economics
Chairman, Departmental Committee on Graduate Studies

Essays on the Macroeconomics of Economic Development

by

Michael Peters

Submitted to the Department of Economics on May 15, 2012, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract

Chapter 1 contains a theory of misallocation. In contrast to a recent literature where misallocation stems from imperfect input markets, I study an economy with non-competitive output markets. This change of focus has two implications. In the cross-section of firms, static misallocation, i.e. the dispersion of marginal products, is not driven by constraints limiting expansion possibilities, but reflects the distribution of mark-ups. Dynamically, the distribution of mark-ups and the economy-wide rate of productivity growth are jointly determined by firms' innovation and entry incentives. The observed cross-country variation in the degree of misallocation might therefore be a symptom of more fundamental differences in the innovation environment. Using firm-level data from Indonesia, I present both reduced form evidence for this mechanism and estimate the models' structural parameters.

In chapter 2, I study the interaction between migration and firms' technology choices. If firms can adapt their production technology, changes in labor supply will induce biased technological adoption. I test this hypothesis using data from one of the largest population transfer programs of the 20th century. After WW2, Germany lost part of its Eastern Territories. Within 2 years, more than 8m people were expelled and transferred to Western Germany. Using individual-level data of the 1960s and 70s I show that refugees experienced substantial reallocation into unskilled occupations, that refugee-rich counties have higher employment shares in occupations which require little formal human capital and that wages in those occupations are especially high in refugee-rich counties.

In chapter 3, which is joint work with Joaquin Blaum and Claire Lelarge, we use a comprehensive dataset of French manufacturing firms to study firms' import behavior. We first document a new fact on the extensive margin of international trade: most firms source only few differentiated varieties of a given product internationally. We then build a simple model based on productivity differences across firms and show that in contrast to the literature on exports, more productive firms do not necessarily source their products from more countries. On the intensive margin, the theory has one robust implication: expenditure shares across varieties should be equalized in the cross-section of firms. While the data is supportive of this prediction across foreign varieties, more productive firms are subject to "home-bias" in that they spend too much on domestic inputs.

Thesis Supervisor: Daron Acemoglu
Title: Charles P. Kindleberger Professor of Applied Economics

Thesis Supervisor: Abhijit Banerjee
Title: Ford International Professor of Economics

Acknowledgements

The gratitude I feel towards my thesis advisors Daron Acemoglu and Abhijit Banerjee is beyond words. In all these years I have immensely benefited from Daron's creativity, clarity and guidance and without his compassionate support, this dissertation would not have been possible. I cannot overstate how influential Daron has been. He will be a constant source of inspiration. I am also deeply indebted to Abhijit. His economic wisdom, his ability to transform an entire paper with one insight and his human warmth, all of which he shared graciously, were essential for the completion of this thesis. Having both of them on my team was both a privilege and a joy.

Moreover, I am extraordinarily thankful to Robert Townsend, whose curiosity and profound intuition shaped my thinking in many long conversations and by being his teaching assistant for more than two years. Most of all however, I thank him for his optimism, his encouragement and his kind words in the summer of 2011 that made all the difference.

While the entire MIT community has been extremely supportive, Marios Angeletos, Iván Werning and in particular Arnaud Costinot deserve special mention. While Marios and Iván only required half a semester each to spark my interest in macroeconomics, Arnaud unfortunately joined MIT too late for me take advantage of his classes in international trade. Both his work and hours of conversation in his office nevertheless left a permanent mark and the third chapter of this thesis would not have been written without him. Finally, I am indebted to my friend and coauthor Stefan Hoderlein, who introduced me to economic research during my undergraduate education. I might not be here without him.

I have also been very fortunate to share these last years with many wonderful classmates and my precious friends Camilo Garcia, Daniel Keniston, Jenny Simon and especially Joaquin Blaum, who has been a faithful friend from day one. Without them, my time at MIT would have been less interesting and far less enjoyable. Similarly, I thank Sascha Boettcher, Christoph Neugebauer, Henning Syllwasschy and Sebastian and Ines Theophil for their support from overseas and for keeping our friendship alive.

I dedicate this dissertation to my parents, whose unconditional love and support brought me here in the first place and who taught me to follow my instincts. I am sincerely grateful for everything they did for me. My biggest gratitude goes to Edna. She selflessly supported my decision to go to MIT and always provided encouragement and perspective - even in times when she needed it most and I did not offer any of it in return. I neither could have nor would have wanted to take this step without her being part of my life.

Contents

Chapter 1

Heterogeneous Mark-Ups and Endogenous Misallocation[1]

1 Introduction

There is a broad consensus in the economic profession that cross-country income differences are to a large degree driven by differences in aggregate productivity (TFP).[2] Aggregative growth theory has traditionally interpreted these disparities as stemming from technological differences across countries. In the last couple of years, a new literature emerged that tends to ignore technological differences and instead focuses on the importance of misallocation of resources across firms within countries to explain the cross-country variation in aggregate TFP. Lately, these contributions have attracted substantial attention, because recent work developed an empirical accounting framework, which allows researchers to measure misallocation using readily available firm-level data. Hsieh and Klenow (2009), following the methodology of Foster, Haltiwanger, and Syverson (2008), show that in an important class of macroeconomic models, firms' productivity is proportional to their marginal products.[3][4] The dispersion of productivity across firms has therefore been interpreted as a key measure of allocative efficiency. And indeed, there is substantial evidence that poor countries are characterized by larger cross-sectional productivity differences, which is consistent with the view that differences in the static efficiency of the resource allocation are important to understand aggregate

[1]I am extremely grateful to my advisors Daron Acemoglu, Abhijit Banerjee and Rob Townsend for their invaluable guidance, encouragement and support. I also thank Ufuk Akcigit, Isaiah Andrews, Thomas Chaney, Maya Eden, Roberto Fattal Jaef, Felipe Iachan, Chang-Tai Hsieh, Daniel Keniston, Amir Kermani, Sam Kortum, Benjamin Moll, Plamen Nenov, Ezra Oberfield, Michael Powell and particularly Marios Angeletos, Arnaud Costinot and Iván Werning as well as seminar participants at MIT, Chicago, Yale, Princeton, Toulouse, LSE, UCLA, Chicago Booth, CREI, Columbia GSB, Brown, Wharton, the RED Mini-Conference on Misallocation and Productivity and the SED. I am also grateful to Ida Fariana from the Badan Pusat Statistik (BPS-Statistics Indonesia) for her cooperation in obtaining the Indonesian micro data, to Monica Martinez-Bravo for providing me with me PODES data and for the financial assistance of the George and Obie Shultz fund to acquire the data necessary for this project. All errors are my own.

[2]I will review the related literature below.

[3]The class of models they consider encompass the Lucas (1978) span-of-control model with heterogeneous firms and an economy with CES demand system, monopolistic competition and factor neutral productivity differences as in Melitz (2003).

[4]The distinction between revenue productivity (given by $TFPR = \frac{py}{l^{1-\alpha}k^{\alpha}}$) and physical productivity ($TFPQ = \frac{y}{l^{1-\alpha}k^{\alpha}}$) is crucial. In the data I only measure $TFPR$ and it is this object, which I will refer to as "productivity". See also Foster, Haltiwanger, and Syverson (2008).

differences in TFP (Hsieh and Klenow, 2009; Bartelsman, Haltiwanger, and Scarpetta, 2009; Collard-Wexler, Asker, and De Loecker, 2011).

In this paper I argue against this sharp conceptual distinction between static misallocation and aggregate technological differences. In particular, in an environment where productivity growth is endogenous and depends on firms' dynamic innovation incentives, the growth rate of aggregate TFP and the static degree of misallocation are jointly determined in equilibrium. Measured productivity differences in the cross-section of firms emerge endogenously and hence are not only informative about the static efficiency of the resource allocation, but also about firms' dynamic incentives, which determine the economy's growth rate. The aggregate implications of within-country productivity dispersion for cross-country income differences might therefore be much larger than previously appreciated.

More specifically, I consider an economy, where imperfect competition on output markets endogenously generates heterogeneous mark-ups across firms. In such an environment, measured productivity differences across producers fully reflect the cross-sectional distribution of prices. These monopolistic distortions cause resources to be statically misallocated, because production factors are not allocated towards the most efficient units but rather to the ones whose monopolistic power is relatively low. From a dynamic point of view, however, these endogenous distortions govern firms' innovation incentives and hence determine the growth rate of aggregate TFP. I study an environment where productivity growth can either be generated by process innovations of current producers or stem from new firms entering the market. While incumbent firms spend innovation resources to increase their mark-up, entering firms reduce mark-ups as they adopt current best-practice techniques and thereby tighten product market competition. A greater equilibrium entry rate is therefore associated with a lower cross-sectional variance of mark-ups and hence less dispersion in measured productivity. Additionally, a higher equilibrium entry rate (usually) also causes the growth rate of aggregate TFP to be higher. Countries with substantial barriers to entry (e.g. license requirements or imperfect markets for venture capital) will therefore both have a high dispersion of productivity as existing firms' monopoly power is unchecked and a low rate of aggregate productivity growth. Conversely, countries that offer a level playing field for new firms will see little productivity dispersion, as entering firms keep monopoly power limited and grow at a faster rate. Hence, higher productivity dispersion not only implies a lower degree of static allocative efficiency, but is also a symptom of a more fundamental difference in the innovation environment, which has direct implications for technological differences across countries.

Modelling productivity differences through monopolistic mark-ups is not merely a new microfoundation, but it gives a new interpretation to the cross-sectional data. In most theories of misallocation, marginal products are not equalized because some firms are constrained in their input choices. This could be due to credit market frictions, differences in non-market access to production factors, or preferential policies, where taxes or the allocation of production licenses are based on firm-specific idiosyncrasies like family ties or political conviction. While these theories stress very different underlying causes of misallocation (and have of course very different policy implications), they share a common economic mechanism why allocative efficiency cannot be achieved: some firms are smaller than they ought to be and smaller than they want to be whenever such constraints are binding. Furthermore, they have the same implications for identifying constrained firms in the data. As constrained producers' marginal products will exceed the one of their unconstrained competitors, high measured productivity will be a sign of the firm facing some binding barrier to expand. According to the theory of this paper, it is not that resources are prevented from flowing to high productivity firms, but it is rather that firms' productivity is high because they can charge a high mark-up. High productivity firms are still smaller than they ought to be, but they are not smaller than they want to be. Hence, high revenue productivity is not a sign of a firm being disadvantaged on some input market, but rather indicative of market power on the output market.

To study these issues, I propose an endogenous growth model in the Schumpeterian tradition, which despite the non-trivial pricing rules generating heterogeneous mark-ups is highly tractable. The cross-sectional distribution of mark-ups has aggregate consequences by reducing TFP and equilibrium factor prices. The model makes precise predictions what aspects of the mark-up distribution matter for aggregate allocations and welfare. In particular, the aggregate economy looks exactly like the single-sector neoclassical growth model with Cobb-Douglas technology and an aggregate TFP term. This aggregate TFP term is the product of two components. While the first one depends only on the distribution of firm-level efficiencies, the second one is entirely determined by the distribution of mark-ups. It is this second term (and only this second term) which is not present in the efficient allocation and is hence a sufficient statistic for the static TFP losses of non-constant mark-ups. Similarly, there is a simple sufficient statistic for the effect of mark-ups on equilibrium factor prices. These two statistics completely summarize the static degree of misallocation given any distribution of mark-ups in the economy.

I then derive the endogenous distribution of mark-ups as an equilibrium object. As mark-ups are determined by firms' relative efficiencies, this distribution depends on the incentives of existing firms to engage in process innovations relative to the benefits for new firms to enter the market. Along the balanced growth path, the distribution of mark-ups takes a Pareto form, whose shape parameter is endogenous. In particular, the shape is equal to the economy's entry intensity, which is given by the equilibrium entry rate relative to the innovation rate of existing firms. If an economy's aggregate productivity growth is largely accounted for by new entrants, the distribution of mark-ups is compressed because entering firms keep monopoly power limited. In such an environment there is little static misallocation of resources. If on the other hand productivity growth is mostly generated by existing producers, the distribution of mark-ups will have a fatter tail, so that productivity differences across firms are pronounced and static allocative efficiency suffers.

I then take the model to the data by analyzing a comprehensive panel data set of manufacturing firms in Indonesia. The empirical analysis has two parts. First, I study the reduced-form implications of the theory. Through the lens of my model, the distribution of mark-ups is identified from firms' revenue productivity. Using this equivalence, I first present evidence that the empirical pattern of productivity is consistent with the mechanism identified in this paper, but less easily reconciled with the above mentioned theories stressing constraints on input choices. In particular, I show that entering firms have low revenue productivity in the cross-section but increasing revenue productivity along their life-cycle. This is consistent with the view that firms manage to charge higher prices as they successfully implement process innovations but harder to square with theories where firms have dynamic incentives to alleviate constraints through saving, political lobbying or gaining access to formerly unavailable inputs. Then I show that the extent of entry into product markets (which I proxy by different geographic regions) reduces both the level and the dispersion of mark-ups as inferred from firms' productivity. I also present evidence that regional variation in the efficiency of the financial system, firm-specific distortions like state ownership or the industrial composition is not causing this pattern.

After establishing those reduced form correlations, I finally estimate the structural parameters of the model. In particular, I estimate the innovation technology and the costs of entry and I allow those parameters to vary at the regional level. With the estimated parameters at hand, I perform two exercises. As the equilibrium distribution of mark-ups is fully determined by the innovation environment and mark-ups are the only reason for firm-level productivity to not be equalized, I can compare the model's prediction with the cross-sectional distribution of productivity observed in the micro data. At the estimated parameters the model can explain 20% of the observed productivity dispersion. Then I perform a simple policy experiment, where I study the implications of a reduction in entry costs. Reducing entry barriers by 10% increases the equilibrium entry rate, decreases the predicted productivity dispersion by 15% and increases welfare by 5-6%. While 10% of these welfare gains are pure static gains in that more intense entry reduces misallocation and hence increases TFP (holding technologies fixed), 30% of the gains are dynamic as improved allocative efficiency induces capital accumulation. The remaining 60% are fully attributed to growth effects as lower entry barriers increase the aggregate growth rate. Hence, this environment is one example, where the direct static gains from reducing misallocation are small relative to the dynamic implications. The strong empirical correlation between within-country productivity differences and aggregate income might therefore be a reflection of fundamental differences, like the costs of entry.

Finally, I consider a brief extension, where I explicitly introduce inter-regional trade. This extension not only shows that allowing for trade leaves the identification of the structural parameters intact, but also generates a new result: along the balanced growth path all regions grow at the same rate because inter-regional terms of trade compensate for differential productivity growth.

Related Literature This paper is mostly related to the literature trying to understand regional productivity differences (Banerjee and Duflo, 2005; Caselli, 2005; Klenow and Rodriguez-Clare, 1997). While a large literature aimed to explain these differences in the context of an aggregate production function,[5] recently various contributions focused explicitly on the behavior of individual firms and their interaction. On the one hand, Hsieh and Klenow (2009) and Restuccia and Rogerson (2008) argue that misallocation of resources across individual firms can have sizable negative effects on aggregate TFP but are agnostic about the reasons for this misallocation. On the other hand, there are papers taking a more structural perspective in that they explicitly model the microstructure of the economy. Acemoglu, Antras, and Helpman (2007) for example argue that contractual incompleteness reduces the incentives to invest into technology and Lagos (2006) shows that in an economy with search frictions, aggregate TFP depends on the primitives of the underlying search technology. Among others, Buera, Kaboski, and Shin (2011), Midrigan and Xu (2010), Jeong and Townsend (2007), Banerjee and Moll (2010) and Moll (2010) consider dynamic models of entrepreneurship and study how capital market imperfections affect aggregate productivity by distorting the allocation of talent across occupations and the allocation of capital across firms. This paper tries to build a bridge between these two approaches. While my model is structural in that firm-level distortions are generated endogenously, the reduced form of my model looks exactly like the economy considered in Hsieh and Klenow (2009) or Restuccia and Rogerson (2008). Hsieh and Klenow (2011) and Fattal Jaef (2011) also stress the dynamic interactions between productivity differences and entry, albeit in a very different way. While they analyze entry incentives given a particular set of distortions firms will eventually face, this paper claims that the causality may go the other way: the equilibrium entry rate will shape the mark-up distribution in the economy as it determines the toughness of competition.

[5]To name a few, Hall and Jones (1999) and Acemoglu, Johnson, and Robinson (2001) argue in favor of the importance of social capital or institutions, Lucas (1988, 1990) stresses the importance of human capital externalities, Parente and Prescott (1994) focus on barriers of technology adoption and Gancia and Zilibotti (2009) identify an important role for the direction of technological progress.

There is also a related literature in the field of international trade stressing the importance of mark-ups. Bernard, Eaton, Jensen, and Kortum (2003) allow for mark-up heterogeneity precisely to account for the positive correlation between measured productivity, firm efficiency and export status, and Atkeson and Burstein (2008) find that variable mark-ups are important to account for the observed pattern of producer and consumer prices. Epifani and Gancia (2011) argue that non-constant mark-ups might be important to assess the welfare effects of trade liberalization - if trade opening reduces mark-ups in sectors whose mark-up was already below average, welfare might decrease as the higher dispersion in mark-ups reduces TFP. Edmond, Midrigan, and Xu (2011) on the other hand provide quantitative evidence from a calibration exercise using plant-level data that the pro-competitive effect of trade-liberalization, whereby the entry of foreign producers decreases mark-ups, induces welfare gains, which are an order of magnitude larger than the usual love-for-variety effects of trade opening. Finally, the importance to distinguish physical productivity from prices has been stressed in the recent empirical literature on how trade liberalization affects firms' productivity. De Loecker (2011) shows that controlling for unobserved prices reduces the estimated productivity increase substantially. Relatedly, De Loecker and Warzynski (forthcoming) find that firms increase their mark-ups upon entry into exporting and argue that mark-ups might be responsible for the large productivity premium of exporting firms usually found in firm-level data. Finally, the distinction between technological progress, aggregate productivity growth and the role of mark-ups has a long tradition in macroeconomics. See for example Basu and Fernald (2002).

In terms of modelling choices, I formalize the evolution of productivity as a Schumpeterian quality ladder model of vertical innovation as in Aghion and Howitt (1992) and Grossman and Helpman (1991). The characterization of the balanced growth path shares some similarities to Klette and Kortum (2004) and Lentz and Mortensen (2008), although I do not consider multi-product firms in my framework. To generate heterogeneous mark-ups, I take an otherwise standard CES demand system, but drop the assumption of different firms producing differentiated varieties.[6] In contrast, different producers engage in Bertrand competition so that prices depend on firms' relative productivities (see also Bernard, Eaton, Jensen, and Kortum (2003) and Acemoglu and Akcigit (forthcoming)).

The structure of the paper is as follows. In the next section I present the model. First, I characterize the static allocations and show that the impact of mark-ups on the aggregate economy is summarized by two sufficient statistics. I then turn to the endogenous determination of equilibrium mark-ups. As mark-ups depend only on the cross-sectional distribution of physical productivity, this requires a model of how relative technologies evolve. I consider a

[6]This is in contrast to for example Foster, Haltiwanger, and Syverson (2008) or Melitz and Ottaviano (2008) who generate heterogeneous mark-ups from a linear demand system with product differentiation.

framework with both entry and productivity improvements by incumbent firms and show that the endogenous stationary distribution of mark-ups is Pareto, where the shape parameter is given by the equilibrium entry intensity. Section three is devoted to the empirical analysis, which proceeds in two steps. In the reduced form analysis, I first assess to what extent different aspects of the productivity data are consistent with my theory of mark-ups and compare the predictions to a simple model with credit constraints. Then I provide evidence that the regional entry intensity is negatively correlated with both the level of mark-ups and their dispersion. Finally, I estimate the structural parameters of the model using GMM, compare the implied mark-up distribution to the firm-level data and consider the welfare implications of a reduction of entry costs in different regions of Indonesia. There I also provide a simple extension of the model, where I explicitly allow for trade between the different regions in Indonesia. Section 4 concludes and an Appendix gathers some of the mathematical details.

2 The Model

Consider the following continuous time environment. There is a measure L of infinitely lived households, supplying their unit time endowment inelastically. There is a unique consumption good and individuals' preferences are given by

$$U = \int_{t=0}^{\infty} e^{-\rho t} ln\left(c\left(t\right)\right) dt, \tag{1}$$

with ρ being the discount rate. The final good, which I take to be the numeraire, is a Cobb-Douglas composite of a continuum of intermediate products. In particular, there is a measure one of differentiated varieties, each of which can be produced by multiple firms. Formally,

$$Y\left(t\right) = exp\left(\int_{0}^{1} ln\left(\sum_{j\epsilon S(\nu,t)} y_j\left(\nu,t\right)\right) d\nu\right), \tag{2}$$

where $y_j\left(\nu,t\right)$ is the quantity of variety ν bought from producer j and $S\left(\nu,t\right)$ denotes the number of firms active in sector ν at time t. Hence, different varieties ν and ν' are imperfect substitutes, whereas there is perfect substitutability between different brands within a variety. The market for intermediate goods is monopolistically competitive, so that firms take aggregate prices as given but compete a la Bertrand with producers offering the same variety. The production of intermediates is conducted by heterogeneous firms and requires capital and labor, both of which are hired on frictionless spot-markets. The only source of heterogeneity across firms is their factor-neutral productivity. In particular, a firm producing variety ν with current productivity[7] q produces output according to

$$f\left(k,l;q\right)=qk^{\alpha}l^{1-\alpha}. \tag{3}$$

Importantly, physical efficiency q is a firm-characteristic, i.e. both within and across varieties, firms differ in that dimension.

Both the set of competing firms $\left[S\left(\nu,t\right)\right]_{\nu}$ and firms' productivities $\left[q_j\left(\nu,t\right)\right]_{j,\nu}$ evolve endogenously through entry of new firms and process innovations by incumbent firms. Existing firms' technology is variety-specific, i.e. once a firm entered in sector ν, it can only produce ν-output. Along their life-cycle, firms can hire workers to increase their productivity $q_j\left(\nu,t\right)$. Alternatively, workers can be employed in an entry sector, where they can try to generate new firms ("blueprints"). Entry efforts are directed, i.e. targeted towards particular sectors. I will specify the details of both the innovation and the entry technology below.

It is useful to characterize the static and dynamic allocations separately. By static allocations, I refer to the equilibrium allocations for a given level of aggregate capital $K\left(t\right)$ and a given number of firms and their productivities $\left[S\left(\nu,t\right),q_j\left(\nu,t\right)\right]$. By construction these allocations neither depend on the inter-temporal preferences of households, nor on the details of the innovation environment. This static analysis will show that the cross-sectional distribution of mark-ups is fully determined from the distribution of physical productivities across firms and characterizes how different moments of this distribution affect the aggregate economy via their effects on mark-ups. To derive the (long-run) equilibrium distribution of mark-ups I therefore need to characterize the evolution of physical productivities. This is done in the dynamic analysis, where I solve for firms' innovation incentives and the equilibrium degree of entry.

2.1 Static Allocations: Heterogeneous Mark-ups and Misallocation

To characterize the static equilibrium, consider first variety ν. Given that production takes place with a constant returns to scale technology, firms compete in prices and different brands of variety ν are perceived as perfect substitutes, in equilibrium only the most productive firm will be active. However, the presence of competing producers (even though they are less efficient) imposes a constraint on the quality leader's price setting. The demand function for intermediaries is given by

$$y(\nu,t)=\frac{Y\left(t\right)}{p\left(\nu,t\right)}, \tag{4}$$

[7] It is unfortunate, that the profession uses the term "productivity" for both the physical productivity (q) and the object measured in the data $\left(\frac{py}{k^{\alpha}l^{1-\alpha}}\right)$. I will sometimes refer to q as "efficiency" or "physical productivity" to be precise about the distinction, but (when clear from the context) I will also refer to q as "productivity".

where $p(\nu,t) \equiv \min_{j\in S(\nu,t)} \{p_j(\nu,t)\}$. As this demand function has unitary elasticity, the most efficient firm has to resort to limit pricing, so that the equilibrium price will be equal to the marginal costs of the second most productive firm, which I will refer to as the follower. Hence,

$$p(\nu,t) = MC(q_F,t) = \frac{\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\left(\frac{R}{\alpha}\right)^{\alpha}}{q_F(\nu,t)} \equiv \frac{\psi(R(t),w(t))}{q_F(\nu,t)}, \tag{5}$$

where $q_F(\nu,t)$ is the follower's productivity.[8] The equilibrium mark-up is therefore given by

$$\xi(\nu,t) \equiv \frac{p(\nu,t)}{MC(q,t)} = \frac{MC(q_F,t)}{MC(q,t)} = \frac{q(\nu,t)}{q_F(\nu,t)}, \tag{6}$$

i.e. a higher quality advantage shields the current producer from competition and allows him to post a higher mark-up. Using (4) and (5) it is easy to derive equilibrium demand as

$$y(\nu,t) = \frac{Y(t)}{\psi(R(t),w(t))}q_F(\nu,t) = \frac{Y(t)}{\psi(R(t),w(t))}\frac{q(\nu,t)}{\xi(\nu,t)}, \tag{7}$$

and the producer's factor demands as

$$k(\nu,t) = \frac{1}{\xi(\nu,t)}\frac{\alpha Y(t)}{R(t)} \text{ and } l(\nu,t) = \frac{1}{\xi(\nu,t)}\frac{(1-\alpha)Y(t)}{w(t)}. \tag{8}$$

Finally, the producing firm's profits are given by

$$\pi(\nu) = (p(\nu) - MC(\nu))\,y(\nu) = \left(1 - \frac{1}{\xi(\nu)}\right)Y. \tag{9}$$

The expressions above show that the cross-sectional variation in profits and factor demands can be entirely traced back to firm-specific mark-ups. Furthermore, it is precisely these varying mark-ups, which induce an inefficient allocation of resources across plants and hence have detrimental effects on aggregate TFP. This is most clearly seen from (8), which shows that

[8] It is at this point where the assumption of the aggregate production function being Cobb-Douglas simplifies the exposition. If the demand elasticity was to exceed unity, the firm might want to set the unconstrained monopoly price in case its productivity advantage over its closest competitor is big enough. To see this, suppose that $Y = \left(\int_0^1 y(\nu)^{\frac{\sigma-1}{\sigma}} d\nu\right)^{\frac{\sigma}{\sigma-1}}$ as in Bernard, Eaton, Jensen, and Kortum (2003). The unconstrained monopoly price is given by $p^M = \frac{\sigma}{\sigma-1}MC = \frac{\sigma}{\sigma-1}\frac{\psi}{q}$, so that the optimal price is given by $p = min\left(\frac{\sigma}{\sigma-1}\frac{\psi}{q}, \frac{\psi}{q_F}\right) = \frac{\psi}{q_F}min\left(\frac{\sigma}{\sigma-1}\frac{q_F}{q}, 1\right)$. Hence, for $q >> q_F$, the firm will not have to use limit pricing. In the limit where $\sigma \to 1$, we get $min\left(\frac{\sigma}{\sigma-1}\frac{q_F}{q}, 1\right) = 1$, which yields (5). We will see below, that the assumption that leading firms will always set the limit price will make the dynamic decision problem of firms very tractable. See also Lentz and Mortensen (2008), who incidentally estimate $\sigma = 1$ using firm-level data from Denmark.

producers' factor demands depend on their mark-up and not on their physical efficiency.

Now consider the implications for measured productivity. In this economy, (revenue) productivity $TFPR$ is given by

$$TFPR(\nu,t) = \frac{p(\nu,t)\,y(\nu,t)}{k(\nu,t)^{\alpha}\,l(\nu,t)^{1-\alpha}} = \psi(R,w)\,\xi(\nu,t)\,. \tag{10}$$

As $\psi(R,w)$ is constant in the cross-section of firms, TFPR is proportional to the mark-up. In the framework of Hsieh and Klenow (2009) and Restuccia and Rogerson (2008), TFPR is proportional to $\frac{(1+\tau_K(\nu,t))^{\alpha}}{1-\tau_Y(\nu,t)}$, where $\tau_K(\nu,t)$ and $\tau_Y(\nu,t)$ are exogenous firm-specific taxes on capital and output. Hence, in this economy firms charging a high mark-up have high productivity and would hence be identified as facing high taxes. This will be important for the empirical analysis.

To characterize the general equilibrium of this economy, let $L_P(t)$ be the number of workers in the production sector.[9] Aggregate output can then be written as

$$Y(t) \;=\; Q(t)\,M(t)\,K(t)^{\alpha}\,L_P(t)^{1-\alpha}\,, \tag{11}$$

where $Q(t) = exp\left(\int_0^1 ln\,(q(\nu,t))\,d\nu\right)$ is the usual CES aggregate of the sectoral productivity levels and $M(t)$ is the *TFP distortion index* given by

$$M(t) \;=\; \frac{exp\left(\int_0^1 ln\left(\xi(\nu,t)^{-1}\right)d\nu\right)}{\int_0^1 \xi(\nu,t)^{-1}\,d\nu} = \frac{exp\left(E\left[ln\left(\xi(\nu,t)^{-1}\right)\right]\right)}{E\left[\xi(\nu,t)^{-1}\right]}. \tag{12}$$

Hence, at the aggregate level, this economy looks exactly like the canonical neoclassical growth model where aggregate TFP is determined both by physical productivity $Q(t)$ and the monopolistic mark-ups summarized in the TFP distortion index $M(t)$. Furthermore, it can be easily verified, that static first-best aggregate output is given by $Y^{FB}(t) = Q(t)\,K(t)^{\alpha}\,L^{1-\alpha}$, so that $Y(t) = Y^{FB}(t)\,M(t)$. $M(t)$ therefore measures precisely the static efficiency losses of monopolistic pricing. Mark-ups also manifest themselves in equilibrium factor prices, as these are given by

$$R(t) \;=\; \frac{\alpha Y(t)}{K(t)}\Lambda(t) = R^{FB}(t)\,M(t)\Lambda(t) \tag{13}$$

$$w(t) \;=\; \frac{(1-\alpha)\,Y(t)}{L_P}\Lambda(t) = w^{FB}(t)\,M(t)\Lambda(t)\,, \tag{14}$$

[9] As workers are used in both the production and the innovation sector, $L_P(t)$ will be determined endogenously. Along the balanced growth path, $L_P(t)$ will be constant. See below.

where $w^{FB}(t)$ and $R^{FB}(t)$ denote the first-best equilibrium factor prices in the absence of monopolistic power and

$$\Lambda(t) = \int_0^1 \xi(\nu, t)^{-1} d\nu = E\left[\xi(\nu, t)^{-1}\right] \tag{15}$$

is the *factor price distortion index.* This term stresses that monopolistic pricing reduces equilibrium factor prices for two reasons.[10] Not only is aggregate TFP (and hence each production factors' marginal product) lower compared to the competitive economy, but even conditional on aggregate TFP, equilibrium factor prices are below their social marginal products as $\frac{R(t)}{MPK(t)} = \frac{w(t)}{MPL(t)} = \Lambda(t)$.[11] Hence, $M(t)$ and $\Lambda(t)$ are sufficient statistics for the static TFP and factor price consequences of monopolistic power. As both $M(t)$ and $\Lambda(t)$ are fully determined from the distribution of mark-ups, Proposition 1 characterizes how this distribution affects aggregate TFP and factor prices in this static economy, i.e. the economy where the distribution of leading productivities $[q(\nu, t)]_{\nu=0}^{1}$ and the level of capital $K(t)$ is given.

Proposition 1. *Consider the static allocations characterized above. Let Z_ξ be a distribution of mark-ups and let $M_Z = \frac{exp(E_Z[ln(\xi^{-1})])}{E_Z[\xi^{-1}]}$ and $\Lambda_Z = E_Z\left[\xi^{-1}\right]$ be the respective distortion indices. Then:*

1. *$M_Z \leq 1$ and $M_Z = 1$ if and only if Z_ξ is degenerate,*
2. *M_Z is homogeneous of degree zero in ξ^{-1},*
3. *A mean preserving spread of $ln(\xi)$ reduces M_Z,*

so that TFP depends on the dispersion of mark-ups. Furthermore,

1. *$\Lambda_Z < 1$,*
2. *$\Lambda_Z M_Z$ is homogeneous of degree one in ξ^{-1},*
3. *A mean preserving spread of $ln(\xi)$ leaves $\Lambda_Z M_Z$ unchanged,*

so that factor prices depend on the level of mark-ups.

[10] This discussion takes the level of capital $K(t)$ as given. Clearly, these distorted factor prices have a third, dynamic effect in that capital accumulation is not efficient. I will come back to this below when I discuss the dynamic allocations.

[11] In particular, equilibrium interest rates will be depressed relative to the interest rate as imputed from the usual growth accounting calibration. From an accounting point of view this is a desirable feature, because in most calibration exercises, the implied interest rates turn out to be counterfactually high (Banerjee and Duflo (2005); Lucas (1990)).

Proof. Follows directly from the definition of M and Λ in (12) and (15). □

Proposition 1 shows that the different moments of the mark-up distribution have very different impacts on the economy. In particular, factor prices are entirely insensitive with respect to a higher dispersion of the underlying mark-up distribution - they only depend on the mean. A shift in the level of mark-ups will therefore reduce factor prices and increase equilibrium profits. For the case of aggregate TFP, exactly the opposite is true: whereas a higher dispersion of mark-ups will show up as a lower TFP for the aggregate economy, TFP is not affected by level shifts. Intuitively, aggregate TFP is determined by the allocation of resources across firms. If all mark-ups increase by a constant proportion, relative prices will be unaffected and hence still provide the appropriate signals about firms' relative productivity. Note that the canonical case of constant mark-ups as generated by a CES demand system with differentiated products is a special case of Proposition 1: TFP will be identical to its efficient competitive counterpart but monopolistic power reduces factor prices.[12]

2.2 Dynamics: The Equilibrium Distribution of Mark-Ups

Proposition 1 shows that different moments of the distribution of mark-ups affect aggregate TFP and equilibrium factor prices differentially. Hence, it is important to know what the equilibrium distribution looks like. Mark-ups are fully determined by relative efficiencies across firms. Misallocation is therefore generated endogenously through competition between producers of different physical productivity. The crucial aspect of the innovation environment is therefore the speed with which leading firms improve their productivity relative to their most advanced potential competitors. If leading firms innovate faster than their potential competitors, they will increase their productivity advantage, which in turn will allow for higher mark-ups. If, on the other hand, less productive firms increase their productivity faster, competition will tighten and static misallocation will be reduced.

To put more structure on the evolution of relative productivities, suppose that firm productivity evolves on a quality-ladder (Aghion and Howitt, 1992; Grossman and Helpman, 1991). Formally, if a firm in sector ν has had $n(\nu,t)$ innovations in the interval $[0,t]$, its productivity is given by $q(\nu,t) = \lambda^{n(\nu,t)}$, where $\lambda > 1$. Hence, innovations are assumed to lead to proportional productivity improvements. This specification of the productivity process is convenient,

[12]Proposition 1 can be seen as the dynamic version of Proposition 1 in Epifani and Gancia (2011, p. 14). While in their static model, *welfare* is homogeneous of degree zero in the level of mark-ups, in my dynamic setting the same holds true for aggregate TFP. Welfare however still depends on the level of mark-ups as dynamic trade-offs will be affected. However, if factor supplies were inelastic, Λ is purely redistributive and M is a sufficient statistic for welfare.

because it implies that monopolistic mark-ups can be expressed as

$$\xi(\nu,t)=\frac{q(\nu,t)}{q_F(\nu,t)}=\frac{\lambda^{n_L(\nu,t)}}{\lambda^{n_F(\nu,t)}}=\lambda^{\Delta(\nu,t)}, \tag{16}$$

where $\Delta(\nu,t)\equiv n_L(\nu,t)-n_F(\nu,t)\geq 0$ is the quality gap between competing firms. Intuitively: firms are able to set high prices if they managed to climb the quality ladder faster than their potential competitors.

In this economy innovations can stem from two sources: either current producers can experience technological improvements, or new firms can enter the market and thereby replace the current quality leader. If a producer with current productivity q experiences an innovation, it reaches the next step of the quality ladder so its new productivity is given by λq. For the case of entry, suppose that leading technologies are common knowledge for the process of innovation so that the entrant in sector ν enters the market with productivity $\lambda q(\nu,t)$, where $q(\nu,t)$ is the leading quality in sector ν. This formulation is appealing in the current context, because it focuses on the different allocational consequences of the two sources of productivity improvements. While both entrants and incumbents increase the frontier technology by the same amount, the implications for equilibrium mark-ups and allocational efficiency are very different. In case the innovation stems from the current producer of variety ν with a quality advantage of $\Delta(\nu,t)$, the equilibrium mark-up for that variety *increases* by a factor λ. However, when productivity growth is induced by entry, the equilibrium mark-up for variety ν *decreases* by a factor $\lambda^{\Delta(\nu,t)-1}$, as the new entrant is only a single step ahead on the quality ladder.[13]

I will show below that along the unique balanced growth path equilibrium, incumbent firms innovate at a constant rate I and each variety ν experiences entry at the constant rate z. However, it is useful to characterize the equilibrium allocations as a function of (I,z) to stress that whatever the particular microfoundation, these two equilibrium outcomes determine aggregate static efficiency by shaping the distribution of productivity gaps and hence mark-ups. In particular, given (I,z), this distribution is stationary and has a closed form representation. The distribution of productivity gaps is fully characterized by the collection $\{\mu(\Delta,t)\}_{\Delta=1}^{\infty}$, where $\mu(\Delta,t)$ denotes the measure of sectors with quality gap Δ at time t. These measures solve the flow equations

$$\dot{\mu}(\Delta,t)=\begin{cases}-(z+I)\,\mu(\Delta,t)+I\mu(\Delta-1,t) & \text{if } \Delta\geq 2\\ -I\mu(1,t)+z(1-\mu(1,t)) & \text{if } \Delta=1\end{cases}. \tag{17}$$

[13]Note that the continuous time formulation of the model precludes the possibility that a variety experiences both entry and a productivity improvement by the current producer, which is of second order.

Intuitively, there are two ways to leave the state (Δ, t): the current producer could have an innovation or there could be entry. The only way to get into this state is by being in state $\Delta - 1$ and then having an innovation. Similarly, all firms in state $(1, t)$ exit this state if they have an innovation and all sectors where entry occurs enter this state. Using (17) it is easy to verify that the unique stationary distribution is given by

$$\mu(\Delta) = \left(\frac{I}{z+I}\right)^{\Delta} \frac{z}{I} \equiv \left(\frac{1}{1+x}\right)^{\Delta} x, \tag{18}$$

where the second equality shows that the mark-up distribution is only dependent on the *entry intensity* defined by

$$x \equiv \frac{z}{I}, \tag{19}$$

which is the ratio of the (endogenous) entry rate and the (endogenous) rate of innovation. This is convenient, because both the cumulative distribution of productivity gaps F_{Δ} and the two distortion indices M and Λ are also constant and fully parametrized by the entry intensity x. In particular, as shown in the Appendix, the distribution of productivity gaps (for a given entry intensity x) is given by

$$F_{\Delta}(d; x) = \sum_{i=1}^{d} \mu(i) = 1 - \left(\frac{1}{x+1}\right)^{d} = 1 - e^{-ln(1+x)d}, \tag{20}$$

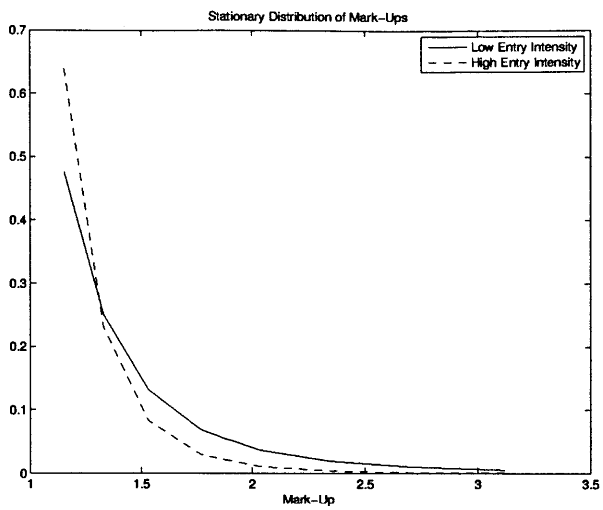
i.e. is exponential with parameter $ln(1+x)$. This also implies that the stationary distribution of mark-ups is Pareto with shape $\frac{ln(1+x)}{ln(\lambda)}$ as[14]

$$F_{\xi}(m; x) = P\left[\lambda^{\Delta} \leq m\right] = F_{\Delta}\left(\frac{ln(m)}{ln(\lambda)}; x\right) = 1 - \left(\frac{1}{m}\right)^{\frac{ln(1+x)}{ln(\lambda)}}. \tag{21}$$

This result is similar to Bernard, Eaton, Jensen, and Kortum (2003), who also generate a (truncated) Pareto distribution of mark-ups in their model. The crucial difference is that here the shape parameter is endogenous as it is determined from the equilibrium entry intensity. If entry is intense, the shape parameter is large so that mark-up heterogeneity and the average mark-up declines.[15] Figure 1 depicts the density of mark-ups for two different values of the entry intensities, which correspond to the inter-quartile range of estimated entry intensities

[14] Of course, Δ is not a continuous variable but only takes integer values. Hence, the distribution functions in (20) and (21) are not differentiable.

[15] In Bernard, Eaton, Jensen, and Kortum (2003), the shape parameter of the derived (unconditional) mark-up distribution is simply the parameter of the underlying Fréchet distribution of physical productivity.



Notes: The figure shows the implied distribution of mark-ups for different values of the entry intensity. They stem from the structural estimation conducted in section 3.3 below and correspond to the observed inter-quartile range of the variation in entry intensities across regions of Indonesia.

Figure 1: Distribution of mark-ups and the entry intensity

across regions of Indonesia (see section 3.3 below). The high-entry environment is characterized by less mark-up dispersion in the cross-section of firms and the average mark-up is lower. Finally, the two distortion indices summarizing the impact of mark-ups on the aggregate economy are also fully determined and given by

$$\Lambda(x) = E\left[\xi^{-1}\right] = E\left[\lambda^{-\Delta}\right] = \frac{x}{(\lambda-1)+\lambda x} \tag{22}$$

$$M(x) = \frac{exp\left(E\left[ln\left(\xi^{-1}\right)\right]\right)}{E\left[\xi^{-1}\right]} = \frac{\lambda^{-E[\Delta]}}{E\left[\lambda^{-\Delta}\right]} = \lambda^{-\frac{x+1}{x}}\frac{(\lambda-1)+\lambda x}{x}. \tag{23}$$

Using (20)-(23) the following Proposition is then immediate.

Proposition 2. *Consider the economy above and let $K(t)$ and $[q(\nu,t)]$ be given. A higher entry intensity x,*

1. *reduces equilibrium mark-ups in a first order dominance sense, as $F_{\xi}(m;x)$ is increasing in x for all m,*

2. *increases aggregate TFP and aggregate income, as $M(x)$ is increasing in x*[16]

[16]To see this, define $s = x^{-1}$ so that $M(s) = \lambda^{-(1+s)}((\lambda-1)s+\lambda)$. $M(s)$ is decreasing in s if

3. increases equilibrium factor prices, as $M(x)\Lambda(x)$ is increasing in x

4. reduces the wedge between factor prices and their marginal product, as $\Lambda(x)$ is increasing in x.

Proposition 2 illustrates the importance of the allocation of innovative resources between entrants and incumbents through the efficiency-enhancing role of entry: in this economy, both types of innovation bring about the same productivity gain but entry reduces factor misallocation by limiting monopolistic pricing. This formalizes the intuition that the static degree of misallocation is high in economies which are characterized by an environment where the entry intensity x is low. The interaction between firm-level distortions and entry incentives has also been discussed in Hsieh and Klenow (2011) and Fattal Jaef (2011) albeit in a very different way. There, a given set of distortions affects equilibrium entry. Proposition 2 shows that the causality might be the other way around: in this economy equilibrium entry shapes the dispersion of productivity differences through its pro-competitive effect.

From this discussion it is also easy to see how one could model the evolution of mark-ups through learning or technology diffusion. Suppose for simplicity that potential frontier technologies were growing at an exogenous rate and call $z+\theta$ the diffusion rate among current producers and z the diffusion rate applying to non-producing firms. If production and the adoption of new techniques are complements, we would expect that $\theta > 0$. If, on the other hand, considerations of vintage capital are important, we would expect $\theta < 0$ as current producers have capital suited to the old technology in place. This model is identical to the one above with $I = z + \theta$, so that $x = \frac{z}{z+\theta}$. Hence, an environment where new firms have a comparative advantage in adopting frontier technologies ($\theta < 0$)) will reduce and compress equilibrium mark-ups and thereby increase static allocative efficiency.

The specific functional form of the stationary mark-up distribution contained in (21) relies of course on the particular entry process. As entrants always enter with the minimal mark-up in the economy, it is intuitive that a higher entry rate will reduce both the level of mark-ups and their dispersion. Here I want to briefly discuss a different assumption concerning the entry process, which suggest that the results in Proposition 2 are more general. Suppose, as before, that incumbent firms always climb one rung of the ladder but that entrants' innovations are larger in that they enter with a blueprint of quality $q(\nu,t)\lambda^{b}$, where $b > 1$. In that environment, the stationary distribution of mark-ups will be exactly the same as before, but just with a higher minimum. In particular, F_{ξ} will be given by (21), where now m takes the

$$\frac{\partial ln\left(M^{BGP}(s)\right)}{\partial s} \equiv h(\lambda,s) = \frac{(\lambda-1)}{(\lambda-1)s+\lambda} - ln(\lambda) < 0.$$

But $h(\lambda,s) < h(\lambda,0) < h(1,0) = 0$.

values $\lambda^b, \lambda^{b+1}, \ldots$ Hence, all the results derived above hold true in that environment. The key is of course the stationary of the distribution. As incumbent firm will always weakly increase their mark-up, the measure of firms charging mark-ups less than λ^b will strictly decrease over time. In the stationary distribution, λ^b is again the minimum mark-up in the economy, which is charged by entering firms. Hence, entry has a pro-competitive effect.[17]

Finally, it is worth noting that the economic channel of these efficiency benefits of entry is very different from the ones stressed in e.g. Barseghyan and DiCecio (2009) or Buera, Kaboski, and Shin (2011). In these papers, inefficient entry is also at the heart of aggregate TFP losses but the mechanism is distinct. Barseghyan and DiCecio (2009) use a model a la Hopenhayn (1992) and show that higher entry costs lower aggregate TFP because it reduces the productivity of the marginal entrant through a general equilibrium effect on factor prices. Similarly, Buera, Kaboski, and Shin (2011) argue that credit market imperfections distort the allocation of talent across sectors if sectors are characterized by different set-up costs. This paper supplements these ideas by showing that lower entry intensities will have negative TFP consequences conditional on the distribution of productivity and capital simply by affecting the stationary distribution of mark-ups and with it allocational efficiency.

2.3 Equilibrium Innovation, Entry and Balanced Growth

The analysis above took innovation and entry rates as given. To microfound those as equilibrium outcomes along the balanced growth path, consider the following innovation and entry technologies. Suppose that if a quality leader with quality advantage Δ wants to achieve an innovation flow rate of I, it has to hire $\Gamma(I, \Delta)$ units of labor. In particular, I assume that

$$\Gamma(I, \Delta) = \lambda^{-\Delta} \frac{1}{\varphi} I^{\gamma}, \tag{25}$$

where φ parametrizes the productivity of the innovation technology and $\gamma > 1$ ensures that the cost function for innovation is convex so that there is a unique solution. The term $\lambda^{-\Delta}$

[17] Alternatively, suppose that new blueprints come in heterogeneous qualities, i.e. with probability $p(j)$ entering firms generate a blueprint of quality $q(\nu, t)\lambda^j$. I show in the Appendix that the stationary distribution of quality gaps Δ (i.e. the analogue of (18)) is given by

$$\mu(\Delta) = \left(\frac{1}{x+1}\right)^{\Delta} \left\{ \sum_{i=1}^{\Delta} p(i)\, x^i \left(\sum_{k=0}^{i-1} \binom{i-1}{k} x^{-k} \right) \right\}, \tag{24}$$

where x is again the entry intensity. The analysis above is a special case of (24), where $p(1) = 1$ and $p(j) = 0$ for $j > 1$. Now suppose that $[p_i]_{i=1}^{\infty}$ decays exponentially, i.e. $p(i) = T\kappa^i$, where T ensures that $\sum_i p(i) = 1$ and $1 > \kappa > 0$. Hence, the higher κ, the higher the chance that conditional on entry, the new blue print is a substantial improvement over the current quality leader which therefore carries a large mark-up. In the Appendix, I show that for different values of κ, a higher entry intensity still induces first-order stochastic dominance shifts in the distribution of mark-ups.

implies that innovations are easier the bigger the productivity advantage Δ and is similar in spirit to the assumption of knowledge capital made in Klette and Kortum (2004) or the setup in Atkeson and Burstein (2010). In particular, this assumption is necessary to make the model consistent with a balanced growth path and with Gibrat's Law, i.e. with the fact the firms' growth rates are independent of size (Sutton, 1997; Luttmer, 2010).[18]

The innovation technology of entrants is assumed to be linear, i.e. each worker can generate a new blueprint with flow rate η per unit of time. Conditional on success, this blueprint can be used in production after hiring χ workers to make the blueprint marketable. I want to think of χ as being affected by policies, i.e. firms have to spend resources to patent the innovations or to get the required licenses (Djankov, La Porta, Lopez-De-Silvaes, and Shleifer, 2002). This is without loss of generality as we can interpret η as containing all the technological requirements to make the blueprint marketable. I will come back to this distinction in Section 3.3 below, when I discuss the identification of the model.

Given this innovation environment, I will now characterize the balanced growth path equilibrium of this economy. The definition of an equilibrium is the usual one.

Definition 3. *Consider the economy above. An equilibrium consists of interest rates and wages* $[r(t), w(t)]_t$*, time paths for consumption and capital* $[C(t), K(t)]_t$*, allocations of capital and employment across producers* $[k(\nu, t|q, \Delta), l(\nu, t|q, \Delta)]_{\nu,t}$*, intermediary good prices* $[p(\nu, t|q, \Delta)]_{\nu,t}$*, value functions* $[V(\nu, t|q, \Delta)]_{\nu,t}$*, entry and innovation rates* $[I(\nu, t|q, \Delta), z(\nu, t|q, \Delta)]_{\nu,t}$ *and aggregate productivity and distortion indices* $[Q(t), M(t), \Lambda(t)]_t$ *such that*

- *prices* $[p(\nu, t|q, \Delta)]$ *and factor demands* $[k(\nu, t|q, \Delta), l(\nu, t|q, \Delta)]$ *are consistent with firms' static profit maximization*
- *aggregate consumption* $[C(t)]$ *is consistent with consumers maximizing utility*
- *entry rates* $[z(\nu, t|q, \Delta)]$ *are consistent with free entry*
- *innovation rates* $[I(\nu, t|q, \Delta)]$ *are optimal given the value functions* $[V(\nu, t|q, \Delta)]$
- *the value functions are equal to the expected present discounted value of profits*
- *the evolution of* $[Q(t), M(t), \Lambda(t)]$ *is consistent with the entry and innovation rates*

[18]Intuitively: per-period profits are given by $\left(1-\lambda^{-\Delta}\right)Y$ and hence concave in Δ. For innovation incentives to be constant, the marginal costs of innovation have to be lower for more advanced firms. The leading term in (25) $\left(\lambda^{-\Delta}\right)$ is exactly the right normalization to balance those effects. Note that firms only generate a high productivity gap when they have multiple innovation *in a row*. Hence, (25) effectively posits that firms can build on their own innovations of the past.

- *prices $[r(t), w(t)]$ are consistent with market clearing, where $r(t) = R(t) - \delta$*

  A balanced growth path equilibrium is an equilibrium where the aggregate growth rate is constant.

In the following, I will focus on equilibria with balanced growth. Along the balanced growth path, aggregate productivity, income, capital and consumption $[Q(t), Y(t), K(t), C(t)]$ grow at constant rates. To characterize the BGP equilibrium, we have to solve for the equilibrium value of a firm, as it is this value function that determines innovation and entry incentives. As per period profits $\pi(\nu, t)$ only depend on the quality gap Δ, physical productivity q is not a state variable for the firm's problem. Hence, let $V(\Delta, t)$ denote the value of a firm with a quality gap Δ at time t. This also implies that entry rates are only dependent on (Δ, t) as neither the costs nor the benefits of entry depend on q. The value function solves the Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{aligned} r(t) V(\Delta,t) - \dot{V}(\Delta,t) &= \pi(\Delta,t) - z(\Delta,t)(V(\Delta,t) - V(-1,t)) + \\ & \max_I \left\{ I(V(\Delta+1,t) - V(\Delta,t)) - w(t)\lambda^{-\Delta}\frac{1}{\varphi}I^{\gamma} \right\}, \qquad (26) \end{aligned}$$

where $z(\Delta, t)$ is the (endogenous) entry rate for a sector with quality gap Δ, which current incumbents take as given. The intuition for (26) is the following: Given the interest rate $r(t)$, the return on the "asset" $(r(t)V(\Delta,t))$ consists of the per-period dividends $\pi(\Delta,t)$ and the asset's appreciation $\dot{V}(\Delta,t)$. Additionally, with flow rate $z(\Delta,t)$ the current leader ceases to be in that position and instead is now one quality step behind the new entrant. Hence, with flow rate $z(\Delta,t)$ a value of $V(\Delta,t) - V(-1,t)$ is destroyed. Similarly, the possibility of investing in technological improvements represents an option value. By spending $w(t)\lambda^{-\Delta}\frac{1}{\varphi}I^{\gamma}$, the firm generates a surplus of $V(\Delta+1,t) - V(\Delta,t)$ with flow rate I. As usual, the occurrence of both a successful incumbent innovation and entry is of second order. As lagging firms neither produce nor innovate, $V(-1,t) = 0$.

In the Appendix I show that the unique solution to (26) is given by

$$V(\Delta,t) = \frac{\pi(\Delta,t) + (\gamma - 1) w(t)\Gamma(I,\Delta)}{\rho + z}, \qquad (27)$$

where I is the optimal innovation rate, which is strictly positive and unique, z is the entry rate, which is constant along the balanced growth path and ρ is the consumers' discount rate. The solution for the value function contained in (27) is intuitive: it is the risk-adjusted net present value of current profits plus the inframarginal rents of the concave innovation technology.

To solve for the equilibrium, note first that given the value function (27), the optimal innovation rate is characterized by the first order condition

$$w(t)\lambda^{-\Delta}\frac{\gamma}{\varphi}I^{\gamma-1} = V(\Delta+1,t)-V(\Delta,t), \tag{28}$$

as $V(\Delta+1,t)-V(\Delta,t)$ is the marginal return of an innovation. The free entry condition is given by

$$w(t) = \eta\left(V(1,t)-\chi w(t)\right), \tag{29}$$

as individuals can either chose to work, earning the wage rate $w(t)$, or they can try to enter with an improved technology. In the latter case, they are successful with a flow rate of η and earn a net value (after paying the entry costs) of $V(1,t)-\chi w(t)$ as they always enter with a productivity advantage of $\Delta = 1$.[19] This is also the reason why equilibrium entry rates are constant across different varieties.

The labor market clearing condition in this economy is then simply given by

$$L \equiv 1 = L_P + L_I + L_E + L_P + \Lambda(x)\frac{1}{\varphi}I^{\gamma} + z\left(\frac{1+\eta\chi}{\eta}\right), \tag{30}$$

where the size of the labor force is normalized to one, L_P is the aggregate amount of production workers, $L_I = \int \lambda^{-\Delta}\frac{1}{\varphi}I^{\gamma}dF_{\Delta} = \Lambda(x)\frac{1}{\varphi}I^{\gamma}$ is the aggregate amount of innovators, $\Lambda(x)$ is the stationary factor price distortion index derived in (22) and $L_E = z\left(\frac{1+\eta\chi}{\eta}\right)$ is the labor requirement to generate an entry rate of z. To determine the equilibrium, note that (upon substituting (27)), (28) and (29) determine the relative wage $\frac{w(t)}{Y(t)}$ as function of the innovation and entry rates (z,I). Together with (14), which showed that the marginal product of labor in the production sector was determined by $\frac{Y(t)}{w(t)} = \frac{1}{1-\alpha}\frac{1}{\Lambda(x)}L_P$, and the labor market clearing condition (30), they fully characterize the equilibrium. In particular, as the two distortion indices $\Lambda(x)$ and $M(x)$ only depend on the entry intensity $x = \frac{z}{I}$, it is useful to focus on (x,I) directly. These are uniquely determined from the two conditions

$$\varphi\frac{1+\eta\chi}{\eta} = \gamma I^{\gamma-1} + \frac{\gamma-1}{\rho+xI}I^{\gamma} \tag{31}$$

$$\varphi = \frac{x\left((1-\alpha)\gamma I^{\gamma-1}(\rho+xI)\frac{\lambda}{\lambda-1} + ((1-\alpha)(\gamma-1)+1)I^{\gamma}\right)}{(\lambda-1)+\lambda x} + \varphi\left(\frac{1+\eta\chi}{\eta}\right) \tag{32}$$

As shown in the Appendix, (31) and (32) not only have a unique solution, but they also imply straight-forward comparative static results. In particular, it can be shown that the

[19] Note that for (29) to be the appropriate free entry condition, I assume that workers are able to insure the idiosyncratic risk of entry. This could be decentralized by a "mutual fund" hiring a continuum of innovators. If there is free entry to open mutual funds, each innovator will be paid the expected return $\eta(V(1,t)-\chi w(t))$.

entry intensity x is decreasing in the entry costs χ and increasing in the efficiency of the entry technology η. Conversely, the innovation rate I is increasing in the efficiency of the innovation technology φ.

To characterize the remaining allocations, note that given the equilibrium innovation rate and entry intensity (x, I), the growth rate of the aggregate productivity index $Q(t)$ is simply

$$g_Q = \frac{\dot{Q}(t)}{Q(t)} = ln(\lambda)(I + z) = ln(\lambda)(1 + x) I, \tag{33}$$

and the two distortion indices $M(x)$ and $\Lambda(x)$ are constant. Given these equilibrium values, the rest of the economy behaves exactly like the neoclassical growth model with exogenous productivity growth - the only difference being that both factor prices and aggregate TFP are affected by $M(x)$ and $\Lambda(x)$, which act like a tax from the agents' point of view. In particular, in the framework of Chari, Kehoe, and McGrattan (2007), $\Lambda(x)$ is exactly the investment and labor wedge and $M(x)$ the efficiency wedge. This is seen from the budget constraint of the representative household[20]

$$\begin{aligned} C(t) + \dot{K}(t) &= w(t) L_P + R(t) K(t) + \Pi(t) - \delta K(t), \\ &= \Lambda(x) MPL(t) L_P + \Lambda(x) MPK(t) K(t) + \Pi(t) - \delta K(t), \end{aligned} \tag{34}$$

where $MPL(t)$ and $MPK(t)$ are the marginal products of labor and capital (which agents take as parametric), L_P is the equilibrium number of production workers and $\Pi(t)$ are the aggregate profits in the economy. Hence, (34) shows that from the consumers' point of view, mark-ups are just like a tax $\Lambda(x)$ on their factor supplies which are rebated back via lump-sum transfers (through $\Pi(t)$). Also, the equilibrium marginal products $MPL(t)$ ($MPK(t)$) are lower than the first-best benchmark by the efficiency wedge $M(t)$ (see (13) and (14)).

Along the balanced growth path, consumption $C(t)$ grows at the constant economy-wide growth rate g_Y, so that normalized consumption $\tilde{c} \equiv \frac{C(t)}{Q(t)^{\frac{1}{1-\alpha}}}$ and capital $\tilde{k} \equiv \frac{K(t)}{Q(t)^{\frac{1}{1-\alpha}}}$ are constant along the BGP. In particular, using the consumer's Euler equation and the equilibrium return to capital $R(t)$ (see (13)), normalized consumption and capital along the BGP are simply

$$\tilde{k} = \left(\frac{\alpha \Lambda(x) M(x)}{\rho + g_Y + \delta}\right)^{\frac{1}{1-\alpha}} L_P(x, I) = \left(\frac{\alpha \lambda^{-\frac{x+1}{x}}}{\rho + g_Y + \delta}\right)^{\frac{1}{1-\alpha}} L_P(x, I) \tag{35}$$

$$\tilde{c} = M(x) \tilde{k}^{\alpha} L_P(x, I)^{1-\alpha} - (\delta + g_Y) \tilde{k}. \tag{36}$$

[20] I show in the Appendix, why (34) is the appropriate budget constraint.

As clearly seen from (35), holding the growth rate fixed, a higher entry intensity x increases the BGP level of capital per production worker as the equilibrium interest rate R will be higher. This is entirely due to the pro-competitive effect of entry: by reducing mark-ups, the demand of capital will be high, which induces capital accumulation. (36) then shows that entry affects consumption even conditional on the level of capital $\tilde{k}$. As $M(x)$ is increasing in the entry-intensity, allocative efficiency will increase. Hence, the composition of productivity growth between entrants and current firms affects economic welfare through four margins: it affects the growth rate g_Q, it governs the dynamic incentives to accumulate capital via $\Lambda(x) M(x)$, it determines static allocative efficiency through $M(x)$ and it controls the allocation of labor.

Proposition 4. *Consider the economy described above. There exists a unique balanced growth path equilibrium. The equilibrium growth rate is given by* $g_Y = \frac{1}{1-\alpha} g_Q$, *where* g_Q *is given in* (33) *and* (x, I) *are uniquely determined from* (31) *and* (32). *The steady-state levels of normalized capital* $\tilde{k}$ *and consumption* $\tilde{c}$ *are given in* (35) *and* (36). *Furthermore, the equilibrium entry intensity* x *is decreasing in the entry costs* χ, *decreasing in the consumers' discount factor* ρ *and increasing in the innovation step-size* λ. *The effect of an increase of the innovation productivity* φ *on the entry intensity* x *is ambiguous.*

Proof. See Appendix. □

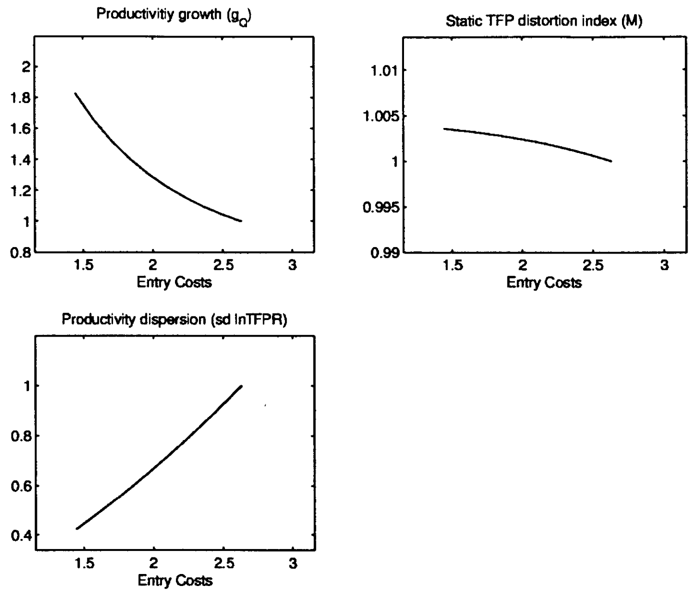
2.4 Static Misallocation, Dynamic Incentives and Aggregate Productivity Differences

According to this paper, the static degree of misallocation and the aggregate productivity growth rate are equilibrium outcomes that are jointly determined. To see this most clearly, note that (using (21) and (33)) we can express the equilibrium distribution of mark-ups (and hence the distribution of measured productivity) as a function of the equilibrium growth rate g_Q and the innovation rate I:

$$F_\xi(m; g_Q, I) = 1 - \left(\frac{1}{m}\right)^{\frac{ln(g_Q/I)}{ln(\lambda)} - 1}. \tag{37}$$

That the cross-sectional dispersion of marginal products[21] and the equilibrium growth rate are related via (37), is particularly important when we try to explain the cross-country variation in aggregate TFP through the cross-firm dispersion of productivity within countries. In particular, once this interdependence is taken into account, the source of variation for the observed productivity dispersion in the cross-country data is crucial. Consider for example

[21] Recall that TFPR, which is equal to the log mark-up, is a geometric average of firms' marginal products.



Notes: The figure shows the comparative statics of the equilibrium growth rate g_Q (Panel A), the TFP distortion index M (Panel B) and the implied standard deviation of log TFPR (Panel C) with respect to the entry costs χ. The parameters stem from the structural estimation conducted in section 3.3 below and I normalize everything to the level of the highest entry costs.

Figure 2: Static Misallocation and Dynamic Incentives: Comparative Statics of Entry Costs (χ)

the case of entry costs χ. In Figure 2 below, I consider the simple comparative static exercise of a reduction of the entry costs. In particular, I calculate the equilibrium outcomes (z, I) for different values of the entry costs and plot the resulting equilibrium growth rate g_Q, the TFP distortion index M and the dispersion of productivity (as measured by the standard deviation of $ln(TFPR)$) relative to the economy with the highest costs. As expected, the equilibrium growth rate and allocative efficiency increase and the dispersion of measured productivity declines, when barriers to entry are lowered.

Now suppose the world in 1900 consisted of a collection of identical economies, which only differed in their entry costs χ. In what sense would productivity differences across firms be informative about the cross-country variation in aggregate TFP? When studying the cross-section of countries in 2011, we would conclude that rich countries show less productivity dispersion than poor countries. In fact, a cross-sectional regression of aggregate TFP on the dispersion of productivity would have an R^2 of one and hence would "explain" the entire cross-country variation in TFP. But if we were to reduce misallocation by equalizing marginal products, we would be disappointed: the implied TFP gains from reallocating resources are small. In fact, at the parameter values used for Figure 2 (which stem from the structural estimation conducted in Section 3.3 below), the static gains are below 1%. More important than the actual number however is the conceptual point that the standard thought experiment of decomposing aggregate TFP into "technology" $(Q(t))$ and "allocative efficiency" $(M(t))$ gets blurred once productivity growth and misallocation are jointly determined. In particular, it is tempting to think of cross-sectional productivity differences as being informative only about the "allocative efficiency"-component, precisely because revenue productivity at the firm-level does not depend on technology (or physical productivity) if marginal products were equalized. This economy is an example, where this conclusion is not warranted. Cross-sectional productivity differences are *very* informative about the "technology" part of aggregate TFP (see (37)), because both of them are equilibrium objects, which are determined simultaneously from firms' dynamic incentives.[22]

3 Empirical Analysis

In this section I will take the model to the data to test its main implications. The empirical analysis has two main parts. In the reduced form analysis in Section 3.2, I use the fact that firms' revenue productivity is proportional to the mark-up (see (10)). Hence, as in Hsieh and Klenow (2009), I measure firm-specific wedges but then interpret the properties of these wedges through the lens of the theory. In particular, I will first present both cross-sectional and life-cycle evidence that the empirical patterns of measured productivity are consistent with this theory but harder to reconcile with theories where firms are constrained in their input choices. Then, I test the main qualitative prediction and show that a higher degree of entry is negatively correlated with the level and the dispersion of mark-ups in a given product market (which I proxy as a geographic region). In Section 3.3 I estimate the model's structural parameters by Generalized Methods of Moments (GMM). Specifically, I will estimate region-specific entry costs (χ_r) and region-specific innovation productivities (φ_r), so that the model generates regional variation in the entry intensity (x_r) as an equilibrium outcome. As the

[22]In fact, there is another example of this economy, which makes this blurred distinction even more precise. Consider a slight variant of the model (the details can be found in the Appendix), where all firms start by using the same technology $q > q_L$ and q_L is an outside technology, which is available to all agents in the economy. Furthermore, suppose that innovation also requires a fixed costs. If both the fixed cost of innovation and the barriers to entry are high enough, there is a unique equilibrium, where this economy is at a steady state and there is no productivity growth. From a static point of view, this economy is first-best efficient, because all marginal products are equalized and there is no cross-sectional productivity dispersion. A reduction in the costs of innovation would cause cross-sectional productivity differences to emerge and allocative efficiency to decline. However, it would also trigger productivity growth and increase both aggregate TFP and welfare. In that world, cross-sectional productivity differences are positively correlated with the efficiency of the innovation environment and will be positively correlated with aggregate TFP in the cross-section of countries.

entry intensity is a sufficient statistic for the entire distribution of mark-ups in the economy, I can then revisit the firm-level data and ask how much of the cross-sectional productivity dispersion the model is able to explain. Furthermore, I will also conduct a policy experiment, where I reduce the costs of entry, and I will decompose the welfare consequences into their static and dynamic components. While static welfare gains are fully described by the change in the TFP distortion index M, the dynamic gains stem from a change in the growth rate, a reallocation of labor across occupations and an increased capital accumulation. These two empirical approaches are not only complementary from an econometric point of view, but they also correspond to two different conceptual exercises. In the reduced form analysis I focus directly on the observed distribution of productivity and use it to infer mark-ups. For the structural estimation I will only employ moments about the dynamic properties of the economy (e.g. the rate of entry and the rate of productivity growth of incumbent firms) but do not use the information contained in the cross-sectional productivity distribution in the set of moments.

For the main part of the empirical analysis I want to think about different regions in Indonesia as representing different replicas of the economy characterized above. The only source of regional variation are the entry costs and the innovation technology. In such a world, the theory implies a concise mapping from regional entry rates to the distribution of productivity of firms active in that particular region. However, Indonesia's different regions are not autarkic closed economies, but there is inter-regional trade. If product markets of Indonesia were fully integrated and each region was producing each of the different varieties, it would be the economy-wide (and not the regional) entry rate that would determine the degree of product market competition In Section 3.4 below, I will therefore consider an environment with trade, which is closer to the spirit of this paper. In particular, I embed my model in a demand structure as in Armington (1969), so that each region produces a particular subset of the product space but all intermediary products are traded. I show that the resulting equilibrium has a very similar structure to the one of the closed economy, so that the reduced form implications remain valid and the structural parameters will be identified from the same moment conditions. However, the extension with trade also implies an important interdependence between regions. While regional physical productivities ($Q(t)$) will grow at different rates, the terms of trade between regions will adjust in a way such that all regions will end up growing at the same rate.

3.1 The Data

I will estimate the model using Indonesian plant-level data. The empirical analysis is based on two data sources, which are described in more detail in the Appendix. My main data set is the Manufacturing Survey of Large and Medium-Sized Firms (Statistik Industri), which for example has also been used in Amiti and Konings (2007) and Blalock, Gertler, and Levine (2008). The Statistik Industri is an annual census of all manufacturing firms in Indonesia with 20 or more employees, and I will use data from 1991 to 2000.[23] The final sample has about 70.000 firms.[24] Most importantly, the data contains information on the geographic region, the respective firm is located in. This allows me to calculate regional entry rates. To do so, I identify a firm as an entrant when it appears first in the data. The regional information is recorded at the level of a regency (kabupaten). There are roughly 240 regencies in Indonesia. As the manufacturing sector in Indonesia is fairly concentrated, many regencies contain too few establishments for a meaningful analysis. I therefore aggregate the data at the level of the 33 provinces.

To be able to control for regional characteristics, I augment this data with information from the Village Potential Statistics (PODES) dataset in 1996. The PODES dataset contains detailed information on all of Indonesia's 65.000 villages. Using the village level data, I then aggregate this information to the province level and match these to the firm-level data. In particular, I exploit information about the financial environment, the state of the infrastructure, the sectoral composition and the density of small, informal firms, which are not in the Statistik Industri data but could of course still exert competitive pressure on the official manufacturing sector. Controlling for these factors should at least alleviate the most pressing concerns about omitted variables, but in the absence of a convincing instrument for the regional entry intensity I will not be able to rule out endogeneity bias entirely for the reduced form analysis.

3.2 Reduced Form Analysis

In the reduced form analysis I will focus on two aspects. First of all I, will present evidence that the empirical pattern of measured productivity is consistent the theory developed here. Then I will test the model's prediction that a higher degree of entry should reduce the level and the dispersion of mark-ups.

[23] As Indonesia experienced a substantial financial crisis in the late 90s, I exclude the data from 1998 (the main year of the crisis) from the main analysis, but I also report the results including this crisis year. The results are qualitatively similar.

[24] To be absolutely precise, the data is collected at the plant level. As more than 90% of the plants report to be single branch entities, I will for the following refer to each plant as a firm. In the context of the model, this distinction is important in that different plants within the same firm are unlikely to compete against each other on product markets.

3.2.1 Productivity and Mark-Ups

According to the theory, firm-specific mark-ups are identified (up to scale) from measured revenue productivity as (see (10))

$$ln\left(TFPR_i\right) = \text{const} + ln\left(\xi_i\right) = \text{const} + ln\left(\lambda\right)\Delta_i. \tag{38}$$

To evaluate if measured productivity is consistent with (38), it is useful to compare the implications with a theory where firms are constrained in their input choices. For the sake of concreteness, I want to focus here on the case of credit-market frictions.[25] Suppose the data was generated by the following economy (which is essentially the setup of Hsieh and Klenow (2009) augmented by collateral constraints). There is a measure one of firms, each of them being the sole provider of a differentiated variety ν. The final good is the usual CES aggregate $Y(t) = \left(\int_0^1 y(\nu)^{\frac{\sigma-1}{\sigma}} d\nu\right)^{\frac{\sigma}{\sigma-1}}$ and each firm produces according to the production function $f(q,k,l) = ql^{1-\alpha}k^{\alpha}$. Capital and labor are traded on a spot market with prices (w, R) and firms face collateral constraints of the form

$$wl + Rk \leq \theta A, \tag{39}$$

where $\theta > 1$ parametrizes the efficiency of the capital market and A denotes firms' wealth. Intuitively: firms can only borrow a multiple $\theta - 1$ of their assets to hire factors of production. In this theory, the two state variables of the firm's problem are given by (q, A), which I assume to be distributed according to some joint distribution G_{QA}. This model has an easy solution, the details of which are provided in the Appendix. In particular, the collateral constraint (39) is binding if and only if

$$\frac{A}{q^{\sigma-1}} \equiv \tilde{A} < \overline{A}\left(w, R, Y, \theta\right), \tag{40}$$

where $\overline{A}(w, R, Y, \theta)$ depends only on parameters and equilibrium quantities, which are common across firms.[26] Hence, firms are constrained if their wealth relative to their productivity q falls below some threshold. Being credit constrained has implications for measured productivity precisely because constrained firms have higher marginal products than their unconstrained

[25] It should be clear below, that I could have chosen another theory, where firms' input choices are constrained. However, theories of credit constraints put enough structure on the patterns we expect to see in the data, that a comparison between the theories is interesting. For example, it is impossible to distinguish this theory to the setup of Hsieh and Klenow (2009), when the distribution of firm-level taxes (τ_Y, τ_K) is unrestricted. In particular, this theory is a special case of their environment with $\tau_K^i = 0$ for all i and $\tau_Y^i = 1 - \xi_i^{-1}$, where ξ_i is firm $i's$ mark-up.

[26] As shown in the Appendix, $\overline{A}(w, R, Y, \theta) = \frac{1}{\theta}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma} Y^{\sigma}\left(\frac{1}{\psi(w,R)}\right)^{\sigma-1}$.

	Dep. Variable: $ln(TFPR)$			
Entry	-0.0381**	-0.0361**	-0.0338**	-0.0350**
	(0.00873)	(0.00876)	(0.00818)	(0.00792)
$ln(\frac{k}{wl})$				-0.162**
				(0.00261)
Year FE	No	Yes	Yes	Yes
Sector FE	No	Yes	Yes	Yes
Region FE	No	No	Yes	Yes
N	58476	58476	58476	58476
R^2	0.000	0.022	0.191	0.248

Notes: Robust standard errors are shown in parentheses. ** and * denotes significance at the 5% and 10% level respectively. "TFPR" is defined by $TFPR = \frac{py}{k^{\alpha}(wl)^{1-\alpha}}$, where α is the sector-specific capital share from the US. "Entry" is an indicator variable if the firm entered in the respective year. $ln\left(\frac{k}{wl}\right)$ is the log of the ratio of firms' capital stock over their wagebill.

Table 1: Entry and $ln\,(TFPR)$

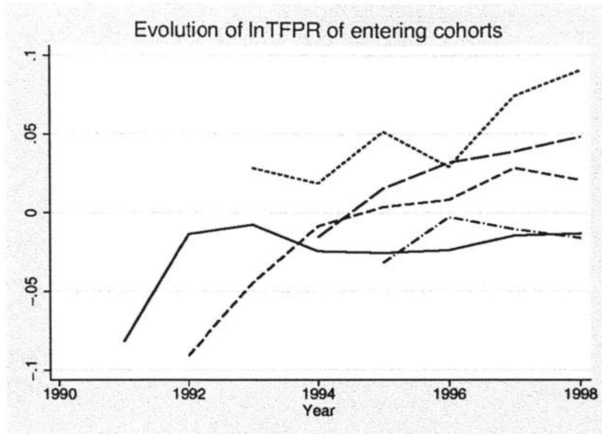
counterparts. It is easy to show that $ln\,(TFPR)$ is given by

$$ln\,(TFPR_i) = \text{const} + \frac{1}{\sigma}\mathbf{1}\left[\tilde{A}_i < \overline{A}\,(w,R,Y,\theta)\right]\left(ln\left(\frac{\overline{A}\,(w,R,Y,\theta)}{\tilde{A}_i}\right)\right), \tag{41}$$

where $\mathbf{1}\,[.]$ is an indicator variable. This shows that (a) constrained firms' revenue productivity is high, (b) the cross-sectional variation is generated by the variation of the single statistic $\tilde{A}$, which measures firms' wealth relative to their productivity and (c) it is the low-$\tilde{A}$ firms that are more likely to be constrained and that have high productivity conditional on being constrained.

To assess if heterogeneous mark-ups are a potential determinant of the firm-level variation of revenue productivity, consider first Table 1, where I compare the productivity of entrants' relative to their incumbent counterparts, by regressing $ln\,(TFPR)$ on a dummy variable if the firm entered in the respective year. Consistent with this paper, entrants' productivity is 3.5% lower than the one of incumbent firms.[27] To square this finding with theories of credit-constraints (or any theory which generates productivity dispersion from constraints on inputs), one would have to argue that entering firms are relatively unconstrained. However, many theories with imperfect capital markets imply that young firms tend to be the constrained agents, that accumulate funds during their life-cycle (see for example Clementi and Hopenhayn (2006)).

[27] A similar result is also found in Foster, Haltiwanger, and Syverson (2008) who observe plant-specific prices and show that young firms do charge lower prices.



Notes: The figure shows the evolution of the average of $ln\,(TFPR)$ for a cohort of firms entering at the respective year after taking out a whole set of year, industry (4 digit) and province fixed effects. "TFPR" is defined by $TFPR = \frac{py}{k^{\alpha}(wl)^{1-\alpha}}$, where α is the sector-specific capital share from the US.

Figure 3: Evolution of $ln\,(TFPR)$ of entering cohorts

It is precisely this life-cycle pattern where the implications of the different theories differ the most. According to this paper, I would expect a cohort of entering firms to increase their revenue productivity as they gradually increase their mark-up over their competitors. Simple models of credit constraints however predict the opposite. If firms are constrained when entering the market, they have incentives to accumulate funds over time so that their relative wealth $\tilde{A}$ will increase. This will reduce their revenue productivity along the life-cycle.[28] Using the panel dimension of the firm-level data I can study this dynamic pattern. In Figure 3, I depict the average (log of) revenue productivity for different cohorts entering the product market in different years. More precisely, the figure shows the evolution of log TFPR after taking out a full set of year, sector and region fixed effects. Figure 3 shows an upward trend in revenue productivity along the life cycle. The same pattern is also reported in Hsieh and Klenow (2011), who note that for firms in India "marginal products of capital and labor (as summarized by TFPR) are also growing with age" (Hsieh and Klenow, 2011, p. 12). According to the theory of this paper, this upward trend is generated by firms increasing their mark-up through productivity improvements. For Figure 3 to be consistent with a model of credit market frictions, entering firms have to become poorer relative to their productivity as they age.

[28] Again, I want to stress that if firms' productivity has a strong trend during the life-cycle (e.g. through learning by doing), it is of course possible for $\tilde{A} = \frac{A}{q^{\sigma-1}}$ to increase in the early stages of the life-cycle. However, there is a strong reason for $\tilde{A}$ to endogenously decline if credit market frictions are important, while mark-ups endogenously increase.

	Dep. Variable: $ln(TFPR)$			
FDI	0.245**			
	(0.0280)			
Foreign loan		0.186**		
		(0.0251)		
Capital market			0.135**	
			(0.0442)	
Capital constrained				-0.0509**
				(0.0176)
$ln(\frac{k}{wl})$	-0.163**	-0.163**	-0.163**	-0.168**
	(0.00261)	(0.00261)	(0.00261)	(0.00635)
Year FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
N	58476	58476	58476	8451
R^2	0.249	0.248	0.248	0.259

Notes: Robust standard errors are shown in parentheses. ** and * denotes significance at the 5% and 10% level respectively. "TFPR" is defined by $TFPR = \frac{py}{k^{\alpha}(wl)^{1-\alpha}}$, where α is the sector-specific capital share from the US. "FDI", "Foreign loan" and "Capital market" are indicator variables if the firm's investment is (partly) financed through FDI, foreign loans or funds from the national capital market. "Capital constrained" is an indicator variable if the firm reported to be capital constrained in the economic census in 1996. $ln\left(\frac{k}{wl}\right)$ is the log of the ratio of firms' capital stock over their wagebill.

Table 2: $ln\,(TFPR)$ and credit constraints

Finally, the Indonesia data allows me to directly look at the importance of credit constraints as it contains information on the source of funding for annual investment expenses. In particular, I observe if firms use FDI, if they attracted foreign loans or if they managed to acquire funds on the national capital market. I want to interpret each of these variables as identifying firms which are relatively less constrained.[29] Columns one to three of Table 2 contain the results for either of these measures. There is a very strong *positive* correlation, which is not what we would expect as unconstrained firms should have low productivity. In column four, I use the cross-section of 1996, where my dataset contains additional information from the economic census, which was added to the usual manufacturing census. Firms were

[29]It might be argued that if a firm raises money on the market, the firm is likely to be constrained - if it had enough wealth to finance everything out of their own funds, it would definitely not be constrained. However, in the context of a developing economy like Indonesia, either of these measures is probably more informative about the availability of funds. Furthermore, they are strongly correlated with firm-size and they apply only to few firms (on the order of magnitude of 1%). See also Blalock, Gertler, and Levine (2008).

asked if the lack of capital was a constraint in their expansion plans. However, column four shows that constrained firms actually have *lower* revenue productivity, which is exactly the opposite of what we expect to find if the variation in revenue productivity was generated from a simple model with credit constraints.

These results are of course not sufficient to reject one of the theories in favor of the other one and this is also not the goal of this exercise.[30] What the results do show however, is that the general patterns we see in the data are at least consistent with a mechanism, that firms with high productivity might *not* be firms that are "unlucky" by being credit constraint or subject to stifling regulation, but that firms have high revenue productivity precisely because they were lucky enough to increase their physical productivity faster than their competitors and are rewarded for doing so with the opportunity to post a high mark-up.

3.2.2 Mark-Ups and Entry

The main qualitative prediction of the theory is that product markets, where entry is intense, should be characterized by low and compressed mark-ups. To study this regional correlation, it is useful to measure mark-ups from firm-specific labor-shares, as (8) implies that[31]

$$\frac{wl}{py}\frac{1}{1-\alpha} \equiv \tau_i = \frac{1}{\xi_i} = \lambda^{-\Delta_i}, \tag{42}$$

where τ_i is observable in the micro- data up to scale.[32] The theory implies that the distribution of mark-ups is Pareto with shape $\frac{ln(1+x)}{ln(\lambda)}$, so that

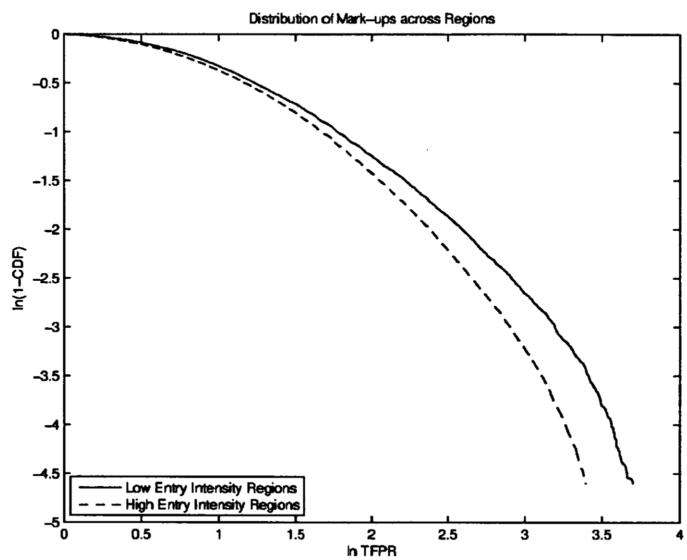
$$ln\left(1 - F_{\xi}\left(m, x\right)\right) = -\frac{ln\left(1+x\right)}{ln\left(\lambda\right)} ln\left(m\right). \tag{43}$$

In Figure 4 below, I plot $ln\left(1 - F_{\xi}\left(m,x\right)\right)$ against $ln\left(m\right)$ and I distinguish regions according to their entry intensity. Figure 4 shows that these distributions are different. In particular, low entry regions display a thicker tail of firms charging a large mark-up. This implication is very much in line with the theory and at the heart of the mechanism of this paper: The distribution of measured productivity is more compressed in high entry environments, as new entrants

[30] Moll (2010) for example presents a model of credit constraints where TFPR is fully determined by firm-productivity. If entering firms have faster productivity growth than incumbents, we would exactly see the pattern depicted in Figure 3.

[31] The reason I employ (42) instead of $ln\left(TFPR\right)$ is that the latter depends on the equilibrium interest rate (see (10)), which are dependent on the regional entry intensity if we think there is a regional capital market. The wagebill in contrast is directly observable in the micro data. Hence, using (42) is theoretically more appealing. In practice, $\rho\left(\Delta\right)$ and $ln\left(TFPR\right)$ are highly correlated so that it does not matter much, which measure I take.

[32] As in Hsieh and Klenow (2009) I take α as the sector-specific capital-share from the US.



Notes: The figure shows $ln\left(1-\hat{F}_{\xi}(m)\right)$, where $\hat{F}_{\xi}(m)$ is the estimated distribution function of log productivity for "high entry" and "low entry" regions. "High entry" regions are those regions with the highest mean entry rates, which cover 25% of the population of firms (dashed line). "Low entry" regions are those regions with the lowest mean entry rates, which cover 25% of the population of firms (solid line).

Figure 4: Distribution of mark-ups in high- and low-entry regions.

keep the environment competitive by keeping prices in line with productivity improvements. Comparing Figure 4 with (43) also shows where the theory is not borne out. The theory implies that both schedules are linear and that the relationship in the entry-intense region should have a steeper slope. While it is the case that the schedule is steeper,[33] the relationship is not linear but shows some curvature.

The relationship between the regional entry rate and the distribution of mark-ups depicted in Figure 4 can of course by purely driven by omitted variables that affect both the entry rate and the measured wedges in the cross-section of firms. To control for at least some of those influences, I will now test the theory in a regression framework. In particular, I will focus on various moments of the distribution of log mark-ups. Using the distribution function for log mark-ups (20), it follows that

$$E\left[ln\left(\xi\right);x\right] = \frac{1+x}{x}, \tag{44}$$

$$sd\left[ln\left(\xi\right);x\right] = ln\left(\lambda\right)\frac{\left(1+x\right)^{1/2}}{x} \tag{45}$$

$$q^{\tau}\left(ln\left(\xi\right);x\right) = -\frac{ln\left(\lambda\right)}{ln\left(1+x\right)}ln\left(1-\tau\right), \tag{46}$$

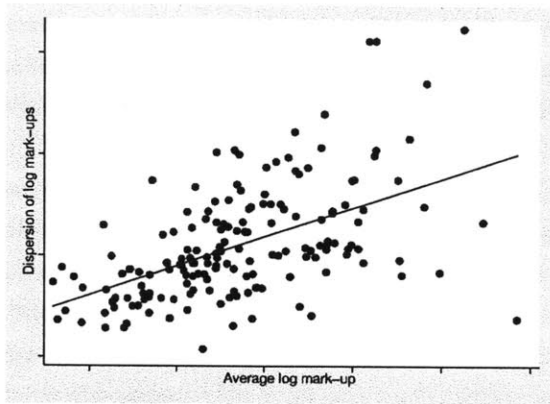
where $E\left[ln\left(\xi\right);x\right]$ and $sd\left[ln\left(\xi\right);x\right]$ denote the mean and standard deviation of log mark-ups given an entry intensity x and $q^{\tau}\left(ln\left(\xi\right);x\right)$ is the τth quantile of the log mark-up distribution. (44) - (46) contain two predictions of the theory. First of all, I expect to find a positive correlation between the average (log) mark-up and the standard deviation of (log) mark-ups. Secondly, the theory implies which regional characteristic is driving this correlation: the entry intensity.

The first prediction is contained in Figure 5, where I depict the simple correlation between the average mark-up and their dispersion. There is a strong positive correlation as the theory predicts.[34] Turning to the second prediction, it is apparent from (44)-(46) that all these moments are decreasing in the entry intensity x. To test this prediction, I consider regressions of the form

$$J_{r,t} = D_t + \phi Entry_{r,t} + IS'_{r,t}\eta + T'_r\beta + u_{r,t}, \tag{47}$$

[33]When I estimate the coefficient on $ln\left(m\right)$ in a simple bivariate regression, the coefficients (standard error) in the high entry region is -1.1494 (0.0018) and -1.051 (0.0036) in the low entry region.

[34]Note that in Figure 5 I take each region-year cell as an observation. The model, which is cast as an economy along its BGP, does not have anything to say about the time-variation in these measures as both are predicted to be constant. However, the variation depicted is driven by the regional variation - when I aggregate everything on the regional level (thereby averaging out the time-variation) there is still a strong positive correlation.



Notes: The figure shows the correlation between the average log mark-up $\overline{ln\left(\tau\right)}_{r,t} = \frac{1}{N_{r,t}}\sum_i ln\left(\tau^i_{r,t}\right)$ and the dispersion of log mark-ups $\hat{\sigma}_{r,t}$, where $\hat{\sigma}^2_{r,t} = \frac{1}{N_{r,t}}\sum_i\left(ln\left(\tau^i_{r,t}\right) - \overline{ln\left(\tau\right)}_{r,t}\right)^2$ and $\tau^i_{r,t} = \frac{w_{r,t}l^i_{r,t}}{(py)^i_{r,t}(1-\alpha)}$, $N_{r,t}$ denotes the number of firms in region r at time t and α is the sector-specific capital share from the US.

Figure 5: Correlation of the average mark-up and mark-up dispersion across regions

where $J_{r,t}$ is one of these moments, D_t is a set of year fixed effects, $Entry_{r,t}$ is the entry rate in region r at time t, $IS_{r,t}$ controls for the industry composition in region r and T_r is a set of regional controls drawn from the regional PODES files described above.[35] The theory predicts that $\phi < 0$. The results are contained in Table 3. Columns one and two take the average log markup as the dependent variable and show that there is a negative correlation with the regional entry rate. While column one presents the simple correlation, column two shows that this correlation is neither driven by a comovement of entry and mark-ups over time nor by regional heterogeneity in their sectoral composition. In column three I include the regional control variables from the PODES files to control for some aspects of regional heterogeneity. Including these leaves the coefficient on the entry rate basically unchanged. While the size of the respective region (when measured by its population) is negatively correlated with the average mark-up, the agricultural employment share and the number of bank branches per village are positive correlated with the average mark-up. Neither the share of villages accessible by asphalted roads (as a measure of infrastructure), nor the number of informal firm seems to have a significant impact. Columns four to eight use different quantiles of the mark-up distribution as the dependent variable. For brevity I report only the specification with the entire set of controls. Again there is a strong positive relation between each of the quantiles and the entry rate, so that more intense entry indeed seems to induce first-order stochastic shifts on the underlying distribution of mark-ups. Note also that the effect becomes weaker,

[35]Note that those regional controls do not vary across time because I only have a single cross-section of the PODES files.

Dep. Variable:	Avg log markup			Quantiles of log markups					Dispersion of log markups	
				95%	90%	80%	75%	50%	std. dev.	IQR
Entry rate z	-0.405**	-0.297**	-0.354**	-0.815**	-0.571**	-0.515**	-0.429*	-0.367*	-0.0795	-0.207
	(0.110)	(0.142)	(0.142)	(0.372)	(0.182)	(0.234)	(0.231)	(0.189)	(0.0863)	(0.226)
ln(population)			-0.0617**	-0.125**	-0.104**	-0.0867**	-0.0786**	-0.0554**	-0.0389**	-0.0245
			(0.0233)	(0.0538)	(0.0360)	(0.0323)	(0.0312)	(0.0234)	(0.0159)	(0.0304)
share of asphalted streets			-0.412	-0.134	-0.537	-0.518	-0.521	-0.622	-0.0987	-0.162
			(0.371)	(0.415)	(0.499)	(0.488)	(0.482)	(0.408)	(0.121)	(0.301)
no of bank branches			0.140*	0.350*	0.254*	0.231**	0.227**	0.143**	0.116**	0.162**
			(0.0704)	(0.202)	(0.125)	(0.0944)	(0.0855)	(0.0630)	(0.0541)	(0.0734)
ln(small firms)			0.0486	-0.0788*	-0.00486	0.0239	0.0462	0.0907*	-0.0426**	-0.0258
			(0.0407)	(0.0400)	(0.0461)	(0.0505)	(0.0517)	(0.0447)	(0.0105)	(0.0285)
agricultural share			0.496*	2.028**	1.177**	0.892**	0.882**	0.397	0.680**	0.794**
			(0.287)	(0.911)	(0.554)	(0.401)	(0.365)	(0.270)	(0.254)	(0.338)
Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industrial Composition	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	168	168	168	168	168	168	168	168	168	168
R^2	0.041	0.596	0.667	0.597	0.638	0.650	0.619	0.626	0.570	0.356

Notes: Standard errors are clustered on the regional level and shown in parentheses. ** and * denotes significance at the 5% and 10% level respectively. The dependent variable are different moments of the distribution of mark-ups $\tau = \frac{wl}{(1-\alpha)py}$, where α is the sector specific capital share from the US. The entry rate $z_{r,t}$ is the share of firms who enter the manufacturing sector in region r at time t. "ln(population)" is the log of the total population in the region in 1996. "share of asphalted streets" is the share of villages in the region, which are accessible by asphalt streets in 1996. "no of bank branches" is the average number of banks per village in the region as of 1996. "ln(small firms)" is the log of the average number of informal firms per village in the region in 1996. "agricultural share" is the average share of the village population whose main income source is agricultural. The industrial composition is controlled for by including the region-year specific value added shares of 2-digit industries in the regression.

Table 3: Entry and the Distribution of Mark-Ups

when I consider smaller quantiles. This is expected from Figure 4, which showed that high and low entry regions differ mostly in the upper tail of the productivity distribution. Finally, in columns nine and ten I use two different measures of dispersion, namely the standard deviation and the interquartile range of log mark-ups. While the point estimate is negative, it is not significant. Hence, the strong positive correlation between the level and the dispersion of mark-ups depicted in Figure 5 is unlikely to stem from the entry intensity as the single regional determinant. However, the structural estimation below will allow me to revisit this result, as I will show that (at the estimated parameters), the implied variation in the dispersion of mark-ups is indeed small.

The model implies that it is only two moments of the underlying distribution of mark-ups, which affect the aggregate outcomes of the economy. In particular, the two moments

$$\Lambda = E\left[\xi^{-1}\right] = E\left[\lambda^{-\Delta}\right] \text{ and } \Psi = exp\left(-E\left[ln\left(\xi\right)\right]\right) = \lambda^{-E[\Delta]} \tag{48}$$

fully determine the impact of mark-ups on the aggregate economy as the two distortion indices $\Lambda = E\left[\lambda^{-\Delta}\right]$ and $M = \frac{\lambda^{-E[\Delta]}}{E[\lambda^{-\Delta}]} = \frac{\Psi}{\Lambda}$ are the two sufficient statistics determining welfare.[36] According to Proposition 2, there is a monotone relation between Λ and Ψ and the entry intensity. To test this prediction, I follow the same approach as in (47) and run regressions of the form

$$\Lambda_{r,t} \quad = \quad D_t + \phi Entry_{r,t} + IS'_{r,t}\eta + T'_r\beta + u_{r,t},$$

where $\Lambda_{r,t}$ is the sample counterparts of (48), and analogously for $\Psi_{r,t}$. The model implies that $\phi > 0$.

The results are reported in Tables 4 and 5. Consider first the case of Λ, contained in Table 4. In the first column I report the simple correlation between the entry rate and $\Lambda_{r,t}$. The coefficient is positive and significant. That this relation is not purely driven by aggregate shocks over time is shown in column two, where I include a full set of year fixed effect. The coefficient hardly changes. Column three controls for the industrial share to account for the fact that some sectors might both be more prone to entry and have lower mark-ups. Like in Table 3 above, I then include the same set of regional controls to at least partially control for some regional characteristics. This does not change the partial correlation with the entry rate, which is not surprising given that none of these regional controls shows a significant correlation with the misallocation measure Λ. In column five, I use the panel structure of the data, which allows me to control for time-invariant regional characteristics by including

[36]Note that Λ is directly identified from the data using (42). Similarly, Ψ can also be directly calculated, as $ln\left(\tau\left(\Delta\right)\right) = -\Delta ln\left(\lambda\right) = -ln\left(\xi\left(\Delta\right)\right)$. Hence, Λ and Ψ are identified from the data regardless of the value of λ.

Dep. Variable:			$\Lambda = E[\lambda^{-\Delta}]$			$ln(\Lambda)$	Λ^{Res}
Entry rate z	0.111**	0.129**	0.112**	0.106**	0.0809**	0.389**	0.0700**
	(0.0260)	(0.0385)	(0.0345)	(0.0300)	(0.0201)	(0.112)	(0.0295)
ln(population)				0.00664		0.0296	-0.00121
				(0.00529)		(0.0200)	(0.00550)
share of asphalted streets				0.0636		0.239	0.00546
				(0.0821)		(0.334)	(0.0644)
no of bank branches				-0.0164		-0.0527	-0.00839
				(0.0165)		(0.0605)	(0.0164)
ln(small firms)				-0.0172		-0.0647	-0.0195*
				(0.0101)		(0.0376)	(0.0101)
agricultural share				-0.00756		-0.0164	-0.0137
				(0.0677)		(0.252)	(0.0696)
Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Industrial Composition	No	No	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	Yes	No	No
N	168	168	168	168	168	168	168
R^2	0.053	0.094	0.537	0.595	0.811	0.560	0.516

Notes: Standard errors are clustered on the regional level and shown in parentheses. ** and * denotes significance at the 5% and 10% level respectively. The dependent variable is the factor price distortion index $\Lambda_{r,t} = \frac{1}{N_{r,t}} \sum_{i=1}^{N_{r,t}} \tau^i_{r,t}$, where $\tau^i_{r,t} = \frac{w_{r,t} l^i_{r,t}}{(py)^i_{r,t}(1-\alpha)}$, $N_{r,t}$ denotes the number of firms in region r at time t and α is the sector specific capital share from the US. The entry rate $z_{r,t}$ is the share of firms who enter the manufacturing sector in region r at time t. "ln(population)" is the log of the total population in the region in 1996. "share of asphalted streets" is the share of villages in the region, which are accessible by asphalt streets in 1996. "no of bank branches" is the average number of banks per village in the region as of 1996. "ln(small firms)" is the log of the average number of informal firms per village in the region in 1996. "agricultural share" is the average share of the village population whose main income source is agricultural. The industrial composition is controlled for by including the region-year specific value added shares of 2-digit industries in the regression. $\Lambda^{Res} = \frac{1}{N_{r,t}} \sum_{i=1}^{N_{r,t}} \tilde{\tau}^i_{r,t}$, where $\tilde{\tau}$ is the residual of a regression of τ on a full set of year and sector fixed effects and the firm-specific capital-labor ratio.

Table 4: Entry and the degree of misallocation: Λ

regional fixed effects. Again this does not affect the coefficient substantially. Columns six and seven contain two additional specifications. In column six, I consider the logarithm of Λ as the dependent variable. The point estimate implies that an increase of the entry rate by one standard deviation (which is equal to 0.1), increases the factor price distortion index by 3%. While this seems very small, note that this coefficient does not have a structural interpretation, because according to the theory it is the entry intensity $x = \frac{z}{I}$, which determines the degree of misallocation. I will come back to the quantitative implications of the theory in section 3.3, when I estimate the structural parameters of the model. Finally, column seven uses a "residual" measure of misallocation. More specifically, instead of constructing Λ directly from the firm-specific labor shares observed in the data, I use the residual labor shares $\tau^{Res} = \tau - \hat{\tau}$, where $\hat{\tau}$ are the predicted values from a regression of τ on a full set of sector fixed effects and the firm-specific capital-labor ratio.[37] Hence, $\Lambda^{Res} = E\left[\tau^{Res}\right]$ is the factor price distortion index after technological heterogeneity at the firm-level is at least crudely controlled for. While the point estimate slightly lower, it is still positive and significant.

The analogous results for $\Psi_{r,t}$ are contained in Table 5. Columns one to four again show a robust positive correlation with the regional entry rate. Contrary to the case of Λ, the regional controls now have some explanatory power but do not change the partial correlation with the entry rate. As Ψ is a non-linear transformation of log mark-ups, the pattern is similar to the one contained in Table 3, i.e. the population size is negatively correlated with mark-ups and both the financial density and the agricultural share is positively correlated with mark-ups. Column five includes a full set or regional fixed effects and columns six and seven again consider the log of Ψ and the residual measure of Ψ as a dependent variable.[38] The correlation is positive and significant (in column seven, the associated p-Value is just above 0.1).

While the empirical results reported above are consistent with the model, I briefly want to consider two alternative explanations. Consider first the case of capital market imperfections, in particular the simple formalization provided above. If the data was generated by that model, I would identify credit constrained firms as charging a high mark-up. Now suppose that the efficiency of the financial system (as parametrized by the collateral multiplier θ) varies across regions so that financially underdeveloped regions will see more financially constrained firms. If the regional entry rate is correlated with the regional financial development, I would conclude that entry reduces monopoly power albeit it is just the case that high entry regions are such that less firms are constrained (or firms are less constrained). As above, I will again consider firms as being relatively unconstrained whenever they are financed through FDI or raise funds

[37]According to the theory, $\frac{k}{l} = \frac{\alpha}{1-\alpha}\frac{w}{R}$ so that the (log of the) capital-labor ratio should be thought of as a control for α.

[38]Note that column 6 is exactly the (negative of the) mean (log) mark-up and is hence numerically identical to column 3 in Table 3.

Dep. Variable:			$\Psi = \lambda^{-E[\Delta]}$			$ln(\Psi)$	Ψ^{Res}
Entry rate z	0.0831**	0.0816**	0.0621*	0.0742**	0.0413**	0.354**	0.183
	(0.0232)	(0.0341)	(0.0305)	(0.0300)	(0.0159)	(0.142)	(0.109)
ln(population)				0.0112**		0.0617**	0.0200
				(0.00443)		(0.0233)	(0.0212)
share of asphalted streets				0.0867		0.412	0.166
				(0.0672)		(0.371)	(0.265)
no of bank branches				-0.0292**		-0.140*	-0.0986
				(0.0134)		(0.0704)	(0.0587)
ln(small firms)				-0.00971		-0.0486	-0.0516
				(0.00793)		(0.0407)	(0.0362)
agricultural share				-0.0983*		-0.496*	-0.485*
				(0.0553)		(0.287)	(0.245)
Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Industrial Composition	No	No	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	No	No	Yes	No	No
N	168	168	168	168	168	168	168
R^2	0.038	0.104	0.650	0.711	0.868	0.667	0.630

Notes: Standard errors are clustered on the regional level and shown in parentheses. ** and * denotes significance at the 5% and 10% level respectively. The dependent variable is $\Psi_{r,t} = exp\left(\frac{1}{N_{r,t}} \sum_{i=1}^{N_{r,t}} ln\left(\tau^i_{r,t,}\right)\right)$, where $\tau^i_{r,t} = \frac{w_{r,t} l^i_{r,t}}{(py)^i_{r,t}(1-\alpha)}$, $N_{r,t}$ denotes the number of firms in region r at time t and α is the sector specific capital share from the US. The entry rate $z_{r,t}$ is the share of firms who enter the manufacturing sector in region r at time t. "ln(population)" is the log of the total population in the region in 1996. "share of asphalted streets" is the share of villages in the region, which are accessible by asphalt streets in 1996. "no of bank branches" is the average number of banks per village in the region as of 1996. "ln(small firms)" is the log of the average number of informal firms per village in the region in 1996. "agricultural share" is the average share of the village population whose main income source is agricultural. The industrial composition is controlled for by including the region-year specific value added shares of 2-digit industries in the regression. $\Psi^{Res} = exp\left(\frac{1}{N_{r,t}} \sum_{i=1}^{N_{r,t}} ln\left(\tilde{\tau}^i_{r,t,}\right)\right)$, where $\tilde{\tau}$ is the residual of a regression of τ on a full set of year and sector fixed effects and the firm-specific capital-labor ratio.

Table 5: Entry and the degree of misallocation: Ψ

Dep. Variable:	Entry rate z			$\Lambda = E[\lambda^{-\Delta}]$	$\Psi = \lambda^{-E[\Delta]}$
Share of firms with foreign loan	0.0608 (0.426)			-0.372 (0.345)	-0.309 (0.209)
Share of firms with FDI		0.381 (0.221)		-0.00561 (0.247)	0.0131 (0.148)
Share of firms with capital market access			1.094 (1.726)	0.431 (0.704)	0.206 (0.484)
Entry rate z				0.107** (0.0274)	0.0741** (0.0280)
Year FE	Yes	Yes	Yes	Yes	Yes
Industrial Composition	Yes	Yes	Yes	Yes	Yes
Regional Controls	Yes	Yes	Yes	Yes	Yes
N	168	168	168	168	168
R^2	0.391	0.398	0.392	0.606	0.719

Notes: Standard errors are clustered on the regional level and shown in parentheses. ** and * denotes significance at the 5% and 10% level respectively. The independent variables are: The share of firms in region r at time t whose investment expenses are at least partially financed through foreign loans, through FDI and through the issuance of equity and/or bonds in the official capital market. The regional controls are the total population, the share of villages, which are accessible by asphalt streets, the average number of banks per village, the average number of informal firms per village and the average share of the village population whose main income source is agricultural in each region in 1996. For the definition of Λ and Ψ see Tables 4 and 5.

Table 6: The importance of credit constraints

on the national capital market of via foreign loans.[39] Given these measures, I investigate whether the share of unconstrained firms is indeed positively correlated with the regional entry rate, and if the entry rate has an independent effect on the degree of misallocation above and beyond regional differences in the share of unconstrained firms. The results are shown in Table 6. Columns one to three show that these measures are positively correlated with the regional entry rate, but insignificantly so. Hence, it is not surprising that columns four and five show that the entry rate is still a significant predictor of the two misallocation measures Λ and Ψ once these measures are controlled for.

The second alternative explanation I want to consider are preferential policies. In particular, the modelling device of "firm-specific taxes" used in Restuccia and Rogerson (2008) or Hsieh and Klenow (2009) is sometimes interpreted as a stand-in for actual policies like regulation, taxes or bureaucratic red tape affecting firms differentially. In particular, firms that I identify as having large monopoly power might in fact just face politically oriented barriers to expand. If additionally regions with good policies are also characterized by more

[39] Recall however the results in Table 2 where I show that these unconstrained firms have relatively high $ln(TFPR)$.

Dep. Variable:	Entry rate z		$\Lambda = E[\lambda^{-\Delta}]$	$\Psi = \lambda^{-E[\Delta]}$
Share of firms	-0.822		-0.239	-0.292
owned by state	(0.660)		(0.180)	(0.183)
Share of firms with		0.0529	0.205**	0.122**
government stake		(0.174)	(0.0837)	(0.0561)
Entry rate z			0.100**	0.0683**
			(0.0299)	(0.0299)
Year FE	Yes	Yes	Yes	Yes
Industrial Composition	Yes	Yes	Yes	Yes
Regional Controls	Yes	Yes	Yes	Yes
N	168	168	168	168
R^2	0.399	0.392	0.618	0.724

Notes: Standard errors are clustered on the regional level and shown in parentheses. ** and * denotes significance at the 5% and 10% level respectively. The independent variables are: The share of firms in region r at time t which are owned by the state or where the local or central government has any financial stake. The regional controls are the total population, the share of villages, which are accessible by asphalt streets, the average number of banks per village, the average number of informal firms per village and the average share of the village population whose main income source is agricultural in each region in 1996. For the definition of Λ and Ψ see Tables 4 and 5.

Table 7: Government ownership

dynamic entry, I would observe exactly the correlations in the data, which I interpret as being informative about the pro-competitive effects of entry. To argue that this is unlikely to be the case, I follow a similar strategy as for the case of credit constraints. In particular, I observe if either the federal or the local government has a financial stake in the firm. In Table 7, I first show that the regional entry rate is uncorrelated with the share of firms with a government stake. Columns three and four then show that the entry rate remains significantly correlated with the two distortion measures Λ and Ψ after these government ownership measures are controlled for. Note however, that the share of firms that have some government financing is also a significant corollary of an improved resource allocation.[40]

[40]Here I briefly discuss the results of various robustness checks, the details of which are available upon request. Concerning the main results reported in Tables 4 and 5, I redid the regressions using the usual robust standard errors instead of the clustered standard errors used in the main analysis, I estimated the regressions using weighted least squares and I included the crisis year 1998. I also redid all the regressions for different cutoffs for the number of firms being present in a region-year cell to be in the final sample. The results are qualitatively unchanged in that all coefficients are positive and mostly significantly so. I also address one specific type of measurement error. In a developing economy like Indonesia, employment of family members is common. While this might be less of a problem for the formal manufacturing firms which I am studying, about 50% of firms do employ unpaid workers (many of which are probably family members) and these unpaid workers constitute on average around 5% of the labor force. If firms do not report these family members in their wagebill, their laborshare will be too low. Hence, I will identify unpaid-worker intensive firms as firms with high monopoly power and if regions with a high entry rate are less reliant on "family firms", I will find the positive relation established above. This is not the case - the occurrence of such firms is uncorrelated with the entry rate and the entry rate remains significantly correlated with the misallocation measures once the

Figure 6: Map of Indonesia

3.3 Structural Estimation

So far I used the model only to qualitatively interpret the distribution of measured wedges, in particular its correlation with the regional entry rate. However, the model is tractable enough to estimate the structural parameters by GMM. As the structural estimation is more data-intensive than the reduced form estimation in Section 3, I partition Indonesia in 7 large regions. These are the islands Sumatra, Sulawesi and Maluku, Kalimantan, Bali and Nusa Tenggara and three regions on Java (the area around Jakarta and Central and West Java), which are depicted in Figure 6 below. For now I want to think of these different regions in Indonesia as being closed economies, but I will consider an extension allowing for inter-regional trade in Section 3.3.3 below.

Table 8 below contains some coarse characteristics of these 7 regions. In the first row I present the regional share of aggregate value added in the Indonesian manufacturing sector. It is clearly seen that Indonesia is very concentrated. Jakarta and the area surrounding it, account for more than 40% of the economic activity and many regions are relatively unimportant. Java as a whole generates 75% of the entire manufacturing value added in Indonesia and account for roughly 80% of producer. There is also evidence of regional specialization. For example more than 60% of all manufacturing firms in Kalimantan are active in the wood industry, while metal and machine industry is mostly located in Java.

Identification To see which parameters of the model are identified from the microdata, consider a single region in a single time period. The model has 8 parameters $(\alpha, \rho, \delta, \varphi, \eta, \chi, \gamma, \lambda)$.

regional shares of such firms are controlled for.

	Sumatra	Jakarta	Central Java	West Java	Bali	Kalimantan	Sulawesi
Share of value added	0.12	0.47	0.12	0.17	0.02	0.08	0.02
Number of Firms	6,428	18,851	13,314	14,457	2,841	1,691	1,900
			Sectoral Composition				
Food, Beverages, Tobacco	0.33	0.17	0.31	0.41	0.22	0.16	0.33
Textile	0.09	0.24	0.24	0.15	0.37	0.01	0.16
Wood Products, Furniture	0.26	0.08	0.12	0.11	0.18	0.62	0.28
Paper Products, Printing	0.03	0.04	0.03	0.04	0.04	0.02	0.03
Chemicals, Petroleum, Coal	0.14	0.14	0.06	0.09	0.01	0.12	0.06
Non-metallic Minerals	0.06	0.17	0.12	0.09	0.12	0.03	0.06
Metals and Machinery	0.10	0.16	0.11	0.11	0.05	0.04	0.06

Notes: The table shows the sectoral composition, the aggregate share of value added and the number of firms of different regions in Indonesia. The sectoral classification is based on the 2-digit ISIC Rev. 2 classification. The sectoral shares refer to the share of firms being active in the respective industry.

Table 8: Economic geography of Indonesia

As I am mostly interested in the innovation environment, I will not be concerned with estimating (α, ρ, δ) but set them exogenously at conventional levels. In particular, I will set the capital share α to 0.3, the depreciation rate δ to 0.1 and the discount rate ρ to 0.05. This leaves 5 parameters to be identified. As the model has a unique equilibrium for the innovation and entry rate, the equilibrium conditions (31) and (32) implicitly define functions

$$z = z(\varphi, \eta, \chi, \gamma, \lambda) \text{ and } I = I(\varphi, \eta, \chi, \gamma, \lambda), \tag{49}$$

which determine the endogenous variables as a function of those parameters. As explained above: I want to estimate the structural parameters without using the cross-sectional productivity distribution explicitly. This will allow me to study how much of the observed productivity dispersion the model can explain and if the model correctly predicts the regional pattern of productivity differences. Hence, I will first focus on three moments, which concern the dynamic outcomes of incumbent firms, entrants and the aggregate economy. In particular, according to the theory

$$\begin{aligned} E\left[g^I\right] &= ln(\lambda) I(\varphi, \eta, \chi, \gamma, \lambda) && (50) \\ E[Entry] &= z(\varphi, \eta, \chi, \gamma, \lambda) \\ g_Q = \frac{\dot{Q}(t)}{Q(t)} &= ln(\lambda)\left(I(\varphi, \eta, \chi, \gamma, \lambda) + z(\varphi, \eta, \chi, \gamma, \lambda)\right), && (51) \end{aligned}$$

were $E\left[g^I\right]$ is the mean growth rate of incumbent firms, $E[Entry]$ is the mean entry rate and g_Q is the growth rate of aggregate productivity. Additionally, I consider the occupational patterns within existing firms, which will be informative about the curvature of the innovation technology γ. Letting $l_I(\Delta)$ be the number of non-production workers working at a firm with mark-up λ^Δ, the non-production share is given by

$$s^I(\Delta) = \frac{l_I(\Delta)}{l_I(\Delta) + l_P(\Delta)} = \frac{1}{\gamma(1-\alpha)\left(\frac{\lambda}{\lambda-1}\frac{\rho+z(\varphi,\eta,\chi,\gamma,\lambda)}{I(\varphi,\eta,\chi,\gamma,\lambda)} + 1\right) + \alpha}. \tag{52}$$

From (50)-(51) and (52) it is apparent that all these moments only depend on the innovation parameters (φ, χ, η) via the resulting equilibrium allocations (z, I). From an estimation point of view this implies that (φ, η, χ) are free parameters, i.e. given some outcomes (z, I), we can always find combinations of such parameters, that implement these as an equilibrium. Hence, from the moment conditions I can directly estimate (z, I) and the parameters (λ, γ) and then use the equilibrium conditions to determine (φ, η, χ) given the estimated (z, I). However, this implies that (φ, η, χ) are not uniquely identified from the microdata - and in fact η and χ

should not be identified according to the theory. To see this, note that the two equilibrium conditions in region r can be written as

$$\varphi_r \tilde{\chi} = \gamma I_r^{\gamma-1} + \frac{\gamma-1}{\rho+z_r} I_r^{\gamma} \tag{53}$$

$$\varphi_r = \frac{z_r \left((1-\alpha)\gamma I_r^{\gamma-1}(\rho+z_r)\frac{\lambda}{\lambda-1} + ((1-\alpha)(\gamma-1)+1) I_r^{\gamma} \right)}{(\lambda-1) I_r + \lambda z_r} + \varphi_r \tilde{\chi} z_r, \tag{54}$$

where $\tilde{\chi} = \frac{1+\eta\chi}{\eta}$. Hence, from the equilibrium conditions, I can only identify the innovation productivity φ and the net entry productivity $\tilde{\chi}$, which has a technological component η governing the efficiency of blueprint creation and an institutional component χ determining the costs of actual entry of new firm once the idea is created. Without more detailed information on where firms spend most of their resources prior to entry, it is not possible to tell η and χ apart. For the policy analysis below I will conceptually think of lowering entry barriers through a reduction of χ, but given that I can only identify $\tilde{\chi}$, this is only an interpretation.[41]

Estimation The moment conditions above hold for all seven regions and at all time period. While (z, I) are region-specific equilibrium outcomes (which are generated through regional variation in the underlying parameters $(\varphi, \tilde{\chi})$), I assume λ and γ to be common across regions. This is useful because it disciplines the model to generate the regional variation in mark-ups only through regional differences in the entry intensity x. Formally, the moment conditions are

$$\begin{aligned}
E\left[g^{I}_{t,t-1,r}\right] &= ln(\lambda) I_r \text{ for all } r = 1,..,R \text{ and } t = 1,...T \\
E\left[Entry_{t,r}\right] &= z_r \text{ for all } r = 1,..,R \text{ and } t = 1,...T \\
E\left[g_{Q,r,t}\right] &= ln(\lambda)(I_r + z_r) \text{ for all } r = 1,..,R \text{ and } t = 1,...T \\
E\left[s^{I}_{i,t}\right] &= \frac{1}{(1-\alpha)\left((\rho+z_r)\frac{\lambda}{\lambda-1}\frac{1}{I_r} + 1\right)\gamma + \alpha} \text{ for } r = 1,...,R \text{ and } t = 1,...,T,
\end{aligned}$$

where $E\left[g_{Q,r,t}\right] = E\left[\frac{Q_r(t+1)-Q_r(t)}{Q_r(t)}\right]$ and $Q_r(t)$ is calculated by aggregating the micro data. Note that these moments hold for all regions *and* all time periods. However, the model (being a balanced growth model) does not have anything to say about the time-variation in the data. Hence, I focus entirely on the regional variation and average all the moments over the time dimension. This yields $4 \times 7 = 28$ moments to estimate 16 parameters $\left(\lambda, \gamma, [z_r, I_r]_{r=1}^{7}\right)$. (53) and (54) then provide the equilibrium mapping to the deep structural parameters $[\varphi_r, \tilde{\chi}_r]_{r=1}^{7}$.

[41] Note however, the evidence reported in Djankov, La Porta, Lopez-De-Silvaes, and Shleifer (2002), which suggests that regulatory barriers are a substantial part of entry costs firms face.

While the first three moments can be directly calculated from the microdata, I do not observe detailed occupational categories in the Indonesian data. Hence, I proxy the share of innovators by the within-firm expenditure share on white-collar non-production workers.

3.3.1 Estimation results

I estimate the parameters via GMM. As the model is overidentified, I choose the efficient GMM estimator,

$$\hat{\theta} = \arg\min_{\theta} \left\{ G(\theta)' \hat{\Omega}^{-1} G(\theta) \right\}, \tag{55}$$

where $\hat{\Omega}$ is the estimated variance-covariance matrix of the 28 moments and G is the matrix of moments. I estimate $\hat{\Omega}$ by bootstrapping the moments from the firm-level data.[42] The estimated parameters are contained in Table 9. The innovation step-size λ is estimated to be equal to 1.17 so that every innovation increases productivity by 17%. This is a little higher than the preferred value of 1.05 in Acemoglu and Akcigit (forthcoming) but within their range of values they consider ($\lambda \in [1.01, 1.2]$). The confidence intervals also show that λ is relatively precisely estimated. The curvature of the innovation technology is estimated to be 1.73 so that the cost function is convex as the theory requires. However, values of $\gamma < 1$ are in the confidence region. The theoretically admissible value of $\gamma = 1$ is the 12% quantile of the distribution for $\hat{\gamma}$.

The last two rows contain the region-specific parameters for the entry and innovation technology. The innovation technology φ_r is estimated to be around 0.2 and the net entry productivity $\tilde{\chi}_r$ around 2.3. This implies that to generate a flow of entry of 10%, the economy requires $\frac{\tilde{\chi}_r}{10} = 0.23$ workers in the entry sector. Similarly, a firm that has just entered and hence has a productivity gap of $\Delta = 1$, requires $\Gamma(1, 0.1) = \frac{1}{\varphi_r} \frac{1}{\lambda} (0.1)^{\gamma} = 0.08$ workers to generate an innovation flow rate of 10%. Hence, at the estimated parameters, incumbent firms seem to enjoy a productivity advantage of generating innovations. Note however, that the confidence intervals for φ_r and χ_r are relatively large.

To see where the variation in the estimated parameter stems from, consider Table 10, where I report the moments in the data and the moments as estimated by the model. The model does a very good job to account for the pattern of entry and the share of non-production workers. This is due to the fact, that the entry rate is precisely estimated (so that the efficient GMM estimator puts a large weight on these moments) and that the share of non-production workers is the only moment that depends on γ, so that γ can basically be chosen freely to match that moment. With respect to the innovation rate of incumbent firms, the model has problems to account for the extreme outcomes in Jakarta and Kalimantan. For Jakarta, we

[42] I use 250 bootstrap replications to estimate $\hat{\Omega}$.

	Sumatra	Jakarta	Central Java	West Java	Bali	Kalimantan	Sulawesi
λ				1.17			
				[1.12;1.27]			
γ				1.73			
				[0.82;2.05]			
φ_r	0.22	0.16	0.07	0.13	0.18	0.18	0.25
	[0.13;0.52]	[0.11;0.39]	[0.05;0.3]	[0.09;0.36]	[0.11;0.46]	[0.11;0.41]	[0.15;0.55]
$\tilde{\chi}_r = \frac{1+\eta_r\chi_r}{\eta_r}$	2.23	2.40	2.63	2.40	2.18	2.76	1.90
	[1.13;5.58]	[1.08;5.95]	[0.89;6.5]	[0.98;6.05]	[0.98;5.56]	[1.41;6.82]	[0.89;4.75]

Notes: 95%-Confidence intervals are shown in brackets. The confidence intervals are calculated using bootstrap. I used 500 bootstrap replications.

Table 9: Results of GMM Estimation

	Growth by Incumbents		Entry Rate		Exp. Share on Nonprod. Workers		Aggregate Growth Rate	
	Data	Model	Data	Model	Data	Model	Data	Model
Sumatra	0.021	0.020	0.148	0.148	0.069	0.070	0.061	0.044
Jakarta	-0.003	0.016	0.140	0.140	0.060	0.058	0.030	0.038
Central Java	0.006	0.007	0.143	0.143	0.025	0.025	0.029	0.029
West Java	0.012	0.012	0.144	0.144	0.044	0.045	0.060	0.035
Bali	0.020	0.016	0.155	0.156	0.051	0.053	0.071	0.040
Kalimantan	0.033	0.019	0.117	0.117	0.078	0.078	0.033	0.038
Sulawesi	0.008	0.020	0.178	0.176	0.071	0.062	0.001	0.048

Notes: The table shows the data moments and the estimated moments for the parameters contained in Table 9.

Table 10: Comparing the implied moments

observe a relatively small entry rate but still aggregate growth. This can only be rationalized through innovation by incumbent firms. For Kalimantan, exactly the opposite is true. Both the entry rate and the growth rate of incumbent firms is high, but aggregate productivity is at the low end. Hence, the model underestimates incumbents' innovation outcomes. Finally, with respect to the aggregate growth rate, the model under-predicts the regional variation, because of the commonality of the step size λ and the restrictions the other moments put on the entry and innovation rate. This is especially seen in the region Sulawesi, which despite its high entry rate does not see any productivity growth. With a constant step size for λ, it is impossible for the model to match these two moments simultaneously. Table 10 also conveys where the respective parameters are identified from. If we compare the regional pattern of the entry rates and the estimated entry costs $\tilde{\chi}$, it is clearly seen that regions with low entry are identified as regions where entry barriers are high. Similarly, regions where the aggregate growth rate and innovation rate of incumbent firms is estimated to be high (Sumatra or Sulawesi), are markets where the innovation technology is of high productivity. Intuitively, level differences of growth mostly identify the innovation productivity φ_r and differences in the entry rate identify the entry costs $\tilde{\chi}_r$.

3.3.2 Mark-ups and the distribution of productivity

Given the estimated parameters, the distribution of mark-ups is therefore fully determined so that I can calculate both the region-specific distortion indices Λ_r and M_r and different measures of productivity dispersion. The results from this exercise are contained in Table 11. Consider first the top panel. The median mark-up is between 10% and 20%, which is consistent with De Loecker and Warzynski (forthcoming) who estimate mark-ups between 17% and 28%.

	Sumatra	Jakarta	Central Java	West Java	Bali	Kalimantan	Sulawesi
Median Mark-up	0.153	0.134	0.076	0.110	0.123	0.178	0.136
Average Mark-up	0.259	0.221	0.119	0.177	0.200	0.309	0.225
Λ	0.757	0.772	0.818	0.791	0.781	0.740	0.771
M	0.983	0.986	0.996	0.991	0.989	0.977	0.986
				Dispersion lnTFPR (std dev)			
Data	0.860	0.751	0.736	0.758	0.769	0.811	0.848
Model	0.201	0.175	0.097	0.143	0.161	0.232	0.178
% Explained	0.233	0.233	0.132	0.189	0.209	0.286	0.210
				Dispersion lnTFPR (75-25)			
Data	1.104	0.990	0.997	1.022	1.092	1.083	1.224
Model	0.226	0.199	0.117	0.165	0.183	0.259	0.202
% Explained	0.205	0.201	0.117	0.161	0.168	0.239	0.165
				Dispersion lnTFPR (90-10)			
Data	2.150	1.905	1.857	1.952	1.966	2.063	2.156
Model	0.452	0.397	0.234	0.330	0.367	0.519	0.403
% Explained	0.210	0.209	0.126	0.169	0.187	0.251	0.187
Correlation of std dev of ln(TFPR))				0.688			
Correlation of iqr of (ln(TFPR))				0.436			
Correlation of q90-10 of (ln(TFPR))				0.705			

Notes: The table contains the implications of the theory for the moments of mark-ups, the distortion indices Λ and M and the dispersion measures of the distribution of $ln\,(TFPR)$. The last three rows contain the regional correlation between the model's prediction and the dispersion measures observed in the data.

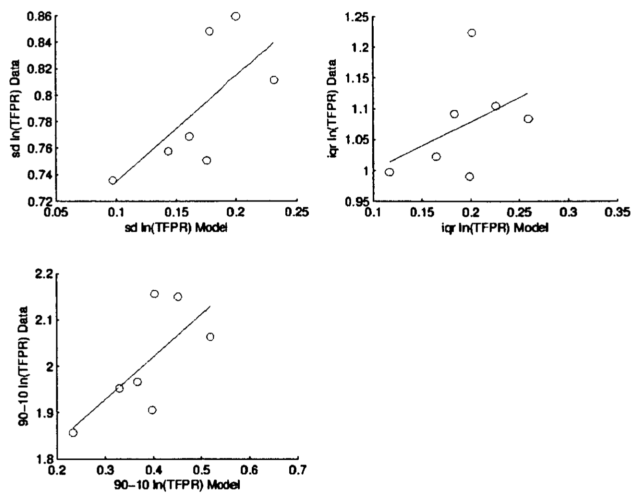
Table 11: Dispersion of TFPR: Data versus Model

The average mark-up is higher, precisely because the theory predicts that the distribution of mark-ups is Pareto. Columns 3 and 4 contain the two distortion indices. The distortions in TFP are modest. On average, heterogeneous mark-ups reduce TFP by about 1.5-2% relative to the first best and there is little variation across regions. This is different for the case of factor prices. Column 5 shows that mark-ups reduce equilibrium factor prices by 25% relative to their social marginal product. The variation across regions is also more substantial than for the case of TFP. The reason for this pattern is contained in Proposition 1: Variation in the entry intensity affects the first moment of the resulting mark-up distribution much more than its second moment. Hence, mark-ups mostly affect factor prices. It is interesting to note that this is exactly what I found in the reduced form analysis, where I showed that the regional entry rate is negatively correlated with the average mark-up but that its relation to the dispersion of mark-ups is insignificant (see Table 3). The same result emerges again from the structural estimation despite the fact that I did not exploit any information about the cross-sectional productivity distribution in the estimation.

In Panels 2 to 4, I compare the implied productivity dispersion (calculated from (44)-(46)) with the one estimated from the micro-data. In the first column of the respective panel I report the estimated standard deviation, inter-quartile range and the 90-10 difference from the Indonesian firm-level data. These numbers are very similar to the one reported in Hsieh and Klenow (2009).[43] The second column reports the implied measures from the theory and third column contains their ratio, which I take as a measure of explanatory power of the theory. On average the model accounts for about 20% of the cross-sectional variation in productivity and this is roughly the same for the different measures. As the variation in the microdata is larger than predicted by the model, the theory does a better job in those regions where the entry intensity is low and the mark-up is high (in Kalimantan for example). The last panel finally contains an important check for the theory: Is the regional variation in the estimated entry intensity correlated with the observed productivity distribution in the microdata? The answer is affirmative: the correlation between the model's implication and the measures observed in the data is between 0.45 and 0.7. A graphical representation of this correlation is contained in Figure 7, which shows a scatter plot of the observed and implied dispersion measures across regions. There is a clear positive relation between the model's prediction and the respective moments in the data, but with only 7 regions, this is of course only suggestive. However, given that the GMM estimation only used information about the dynamic environment, there is no mechanical reason for this correlation to occur.

Another way to judge in how far this mechanism is successful to account for the cross-

[43]They find that the respective numbers are 0.74, 0.97 and 1.87 for the case of China and 0.69, 0.79 and 1.73 for the case of India. See Hsieh and Klenow (2009, Table 2, p. 1418).



Notes: The figure shows a scatter plot of the model's prediction and the dispersion measured observed in the data. It displays the data contained in Table 11

Figure 7: Model vs. Data: Correlation of dispersion of $ln\left(TFPR\right)$

sectional productivity dispersion across firms is to directly include this information in the set of moments. Specifically, let me add the within-region variance of log productivity to the set of moments utilized so far. Using (44), these additional moment conditions are given by

$$E\left[\left(ln\left(TFPR_{i,t,r}\right)-E\left[ln\left(TFPR_{i,t,r}\right)\right]\right)^{2}\right] = ln\left(\lambda\right)\frac{1+x_{r}}{x_{r}^{2}}=ln\left(\lambda\right)^{2}\left(\frac{I_{r}}{z_{r}}\right)^{2}\left(\frac{I_{r}+z_{r}}{I_{r}}\right) \tag{56}$$

which again hold for all regions and all time periods. Like before, I will only use the regional variation and average out the time dimension. By estimating the parameters and studying the implied moments, it is apparent which margins of the data are putting discipline on the theory to account for both the cross-sectional dispersion and the dynamic behavior of the economy simultaneously. In Table 12, I again report the comparison between the data and the model's implication at the estimated parameters. In the last two columns, I show that the model is able to account for the cross-sectional dispersion rather well. However, this comes at the expense to account for other features of the data. From (56) it is apparent that in order to generate a large cross-sectional dispersion of productivity, the theory requires the step size λ to be large and the entry intensity x to be small. These conflicting objectives are seen in Table 12. At the estimated parameters, the model underestimates the entry rate, generates too large a growth rate for incumbent firms and implies a rate of aggregate productivity growth, which exceeds the one observed in the data. Furthermore, the implied median mark-up is around

	Growth by Incumbents		Entry Rate		Exp. Share on Nonprod. Workers		Aggregate Growth Rate		variance $ln(TFPR)$	
	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
Sumatra	0.021	0.058	0.148	0.079	0.069	0.067	0.061	0.079	0.739	0.739
Jakarta	-0.003	0.044	0.140	0.070	0.060	0.057	0.030	0.063	0.565	0.565
Central Java	0.006	0.026	0.143	0.043	0.025	0.047	0.029	0.038	0.534	0.534
West Java	0.012	0.042	0.144	0.066	0.044	0.057	0.060	0.060	0.570	0.570
Bali	0.020	0.048	0.155	0.075	0.051	0.060	0.071	0.069	0.590	0.587
Kalimantan	0.033	0.047	0.117	0.069	0.078	0.061	0.033	0.066	0.655	0.654
Sulawesi	0.008	0.050	0.178	0.069	0.071	0.064	0.001	0.069	0.723	0.724

Notes: The table shows the data moments and the estimated moments when I include the moment condition for the cross-sectional variance of $ln(TFPR)$ (56).

Table 12: Moments using the cross-sectional information

70%, which is also counterfactually high. Hence, the theoretical mechanism laid out here is not able to account for entire productivity dispersion in the cross-section of firms - according to Table 11 roughly 80% of it remains unexplained.

3.3.3 Policy and Welfare

One advantage of a theory that creates misallocation endogenously is that we can use it to think about policy and the welfare implications thereof. Here I want to think about one particular policy change, namely a reduction in the entry costs χ. Not only is this the only parameter in the model which can reasonably be affected by policy, but it also affects the precise margin, which the theory stresses: the entry intensity. Such a change will affect the equilibrium outcomes (z, I) and hence both the entry intensity $x = \frac{z}{I}$ and the equilibrium growth rate $g_Q = ln(\lambda)(I + z)$. To trace the effect on the equilibrium allocations, recall that the BGP levels of (normalized) capital and consumption were given by

$$\begin{aligned}
\tilde{k}(x, I, g_Q) &= \left(\frac{\alpha\Lambda(x) M(x)}{\rho + \frac{1}{1-\alpha} g_Q + \delta}\right)^{\frac{1}{1-\alpha}} L_P(x, I) \\
\tilde{c}\left(x, g_Q, \tilde{k}, L\right) &= M(x) \tilde{k}^{\alpha} L_P(x, I)^{1-\alpha} - \left(\delta + \frac{1}{1-\alpha} g_Q\right) \tilde{k}. \qquad (57)
\end{aligned}$$

Let us denote the variables after the policy change by "*". Then we can decompose the change in consumption in the three terms as follows:

$$\frac{\tilde{c}^*}{\tilde{c}} = \underbrace{\underbrace{\frac{\tilde{c}\left(x^*, g_Q, \tilde{k}(x, I, g_Q), L\right)}{\tilde{c}\left(x, g_Q, \tilde{k}(x, I, g_Q), L\right)}}_{\text{TFP}} \underbrace{\frac{\tilde{c}\left(x^*, g_Q, \tilde{k}(x^*, I^*, g_Q), L^*\right)}{\tilde{c}\left(x^*, g_Q, \tilde{k}(x, I, g_Q), L\right)}}_{\text{Reallocation}}}_{\text{Misallocation}} \underbrace{\frac{\tilde{c}\left(x^*, g_Q^*, \tilde{k}\left(x^*, I^*, g_Q^*\right), L^*\right)}{\tilde{c}\left(x^*, g_Q, \tilde{k}(x^*, I^*, g_Q), L^*\right)}}_{\text{Growth}}. \qquad (58)$$

The TFP-component of (58) holds the growth rate, the capital stock and the allocation of labor constant and contains only the static gains of entrants fostering competition and thereby increasing aggregate TFP through a reduction of static misallocation. The Reallocation-Component contains the change in consumption, which is due to the fact that a change in the entry intensity changes factor-prices and hence the return to capital and the allocation of labor. The sum of those effects are what I call the Misallocation-Component of entry, because they comprise the change in consumption holding the growth rate (counterfactually) fixed. The Growth-Component captures this last dynamic effect as changes in the growth rate will affect both the level of (normalized) consumption and capital. Before presenting the results let me stress that (58) decomposes the change in *steady-state* consumption. My model economy however has non-trivial transitional dynamics. Not only are the usual neoclassical dynamics present due to the accumulation of capital, but the distribution of mark-ups will also not jump to its new stationary distribution but its behavior will be governed by the flow equations given in (17). Furthermore, as long as the distribution of mark-ups is not stationary, the problem firms face is not stationary either so that both entry and innovation rates will not be constant. This makes the characterization intractable.

The results of the policy experiment are contained in Table 13. The first row contains the size of the policy change, namely a decrease in the entry costs by 10%. This increases the entry-intensity by about 25% and reduces the dispersion of productivity by about 15%. In terms of the misallocation measures Λ and M, this increase in the entry intensity increases factor prices (relative to their marginal product) by 1% to 2.5% but has a small effect on static misallocation (as measured by M), where the increase is at most half a percentage point.[44] Columns 6 to 8 trace the effect of this policy change on the equilibrium allocation of workers across occupations. While the number of workers in the entry sector, increases by roughly 2-3%, these workers are mainly drawn from incumbent firms who see their innovation employment decline. In fact, in most regions the reduction of entry costs frees up labor resources and production employment increases. However, this is not necessarily so. In Central Java, whose estimated innovation productivity is low, the change in factor prices causes formerly employed production workers to start working in the entry sector. Finally, a reduction of the entry costs increases the equilibrium growth rate in all regions by about one or two tenth of a percentage point.

The second panel shows the decomposition contained in (58). The reduction in static misallocation raises the level of consumption by around half a percent. The Reallocation-component is larger, as more intense entry increases the number of production workers and the level of capital (except for Central Java). On average, an improved allocation of resources through a change in entry incentives increases consumption by around 2%. Finally, the Growth-component induces a slight fall in those consumption gains, because an increase in the growth rate will decrease the level of consumption as more resources are needed to have the capital stock grow at the economy-wide growth rate. The next panel contains the same decomposition for the change in the capital stock. By construction there is no "static" gain

[44]There are two reasons why this number is so much smaller than the one obtained in Hsieh and Klenow (2009). First of all, the "liberalization" experiment conducted here is much smaller than going from "India to the US". As the mechanism stressed here can explain 20% of the cross-sectional dispersion of productivity, the change in entry costs considered here reduces such dispersion only by 3% (15% times 20%). Secondly, the TFP aggregator in my economy takes the Cobb-Douglas form (i.e. $\sigma = 1$), whereas Hsieh and Klenow (2009) consider a CES demand system with $\sigma = 3$. The TFP gains from a reduction of distortions however are increasing in the elasticity of substitution.

Change in:	Sumatra	Jakarta	Central Java	West Java	Bali	Kalimantan	Sulawesi
entry costs $(\frac{\chi^*-\chi}{\chi})$	-10.00	-10.00	-10.00	-10.00	-10.00	-10.00	-10.00
entry intensity $(\frac{x^*-x}{x})$	24.46	25.92	30.33	27.53	26.23	24.18	24.88
productivity dispersion $(\frac{\sigma^*_{\ln TFPR}-\sigma_{\ln TFPR}}{\sigma_{\ln TFPR}})$	-14.52	-14.78	-14.72	-14.82	-14.63	-14.83	-14.34
Λ $(\frac{\Lambda^*-\Lambda}{\Lambda})$	2.25	1.98	0.96	1.58	1.79	2.65	1.96
M $(\frac{M^*-M}{M})$	0.46	0.36	0.12	0.25	0.30	0.61	0.36
labor employed by entrants $(\frac{L^*_E-L_E}{L_E})$	2.01	2.60	3.91	3.11	2.59	2.11	2.01
non-production workers $(\frac{L^*_I-L_I}{L_I})$	-13.02	-14.13	-18.12	-15.60	-14.66	-12.19	-13.77
production workers $(\frac{L^*_P-L_P}{L_P})$	1.06	0.62	-1.07	0.08	0.51	1.16	0.92
growth rate $(g^*_Q - g_Q)$	0.13	0.16	0.28	0.21	0.20	0.08	0.19
			Change in BGP consumption				
TFP-Component	0.55	0.44	0.14	0.30	0.37	0.73	0.44
+ Reallocation-Component	2.21	1.56	-0.71	0.76	1.32	2.64	1.85
+ Growth-Component	1.76	0.99	-1.73	0.01	0.65	2.34	1.25
			Change in BGP capital				
Reallocation-Component	5.00	4.02	0.46	2.71	3.53	5.92	4.29
+ Growth-Component	3.66	2.35	-2.43	0.54	1.57	5.05	2.50
			Change in Welfare				
Full Welfare Gains	5.74	5.79	6.35	6.23	6.46	4.82	6.79

Notes: The table contains the model's implications of a reduction of entry costs within regions of Indonesia. All entries are in percentage terms. The details are contained in the main body of the text.

Table 13: Effects of a reduction of barriers to entry

where factors are held constant. Given the changes in M and Λ (especially the latter) and the reallocation of workers, normalized capital increases by around 3-4%, with again Central Java being the exception. As for the case of consumption, the higher growth rate compensates slightly for those improvements in the allocation of resources, which reduces the normalized capital-stock.

These results concern the *allocations* and *not welfare.* The reason is of course that welfare is directly affected by the growth rate. In particular, along the balanced growth path, welfare is given by

$$\begin{aligned} W(g_Q, x, I) &= \int_{t=0}^{\infty} e^{-\rho t} ln\left(C\left(t\right)\right) = \int_{t=0}^{\infty} e^{-\rho t} ln\left(\tilde{c} Q\left(0\right)^{\frac{1}{1-\alpha}} e^{\frac{g_Q}{1-\alpha} t}\right) dt \\ &= \frac{ln\left(Q\left(0\right)^{\frac{1}{1-\alpha}}\right)}{\rho} + \frac{ln\left(\tilde{c}\left(x, g_Q, \tilde{k}\left(x, I, g_Q\right)\right)\right)}{\rho} + \frac{g_Q}{\left(1-\alpha\right)\rho^2}. \end{aligned} \tag{59}$$

As $Q(0)$ is an initial condition, I measure changes in welfare by their implied consumption changes holding the growth rate constant, i.e.

$$dc\left(\left[g_Q^*, x^*, I^*\right], \left[g_Q, x, I\right]\right) = exp\left(\rho\left(W\left(g_Q^*, x^*, I^*\right) - W\left(g_Q, x, I\right)\right)\right)$$

is the change in consumption generating a change in welfare $W\left(g_Q^*, x^*\right) - W\left(g_Q, x\right)$ if the growth rate was constant. The results of that exercise are contained in the last panel of Table 13. In particular, the small increases in the growth rate have large effects on welfare - at least when viewed through equation (59), which contains welfare along the balanced growth path. The induced welfare gains are at the order of magnitude 5-6% of total consumption. However, these welfare gains are surely a large upper bound. During the transition, I suspect the growth rate to be lower and the now higher capital stock has to be accumulated. Hence, the higher growth rate is only achieved in the future (which is of course discounted at rate ρ) and households have to forego consumption to save. The welfare calculation in Table 13 does not take into account either of these concerns.

3.4 Inter-regional trade

Finally, I want to briefly extend the environment to allow for inter-regional trade. This is important for two reasons. First of all, this extension will show that trade across regions of Indonesia will not necessarily invalidate the empirical exercise above. In particular, this section shows that introducing trade in a way akin to Armington (1969) leaves the structure of the equilibrium essentially unchanged. This implies that all the predictions for the reduced form analysis still hold true and that the moment conditions used in the structural estimation still identify the structural parameters. Hence, the assumption of the different islands being autarkic economies is not essential (although it, of course, depends on how inter-regional trade is exactly modelled). Secondly, this extension is important when thinking about (long-run) TFP differences across countries. Above I suggested that the differences in productivity dispersion might to a large degree be a symptom of, for example, institutional differences that have implications for the growth rate of aggregate productivity. Does this imply that such (small) differences in the entry environment will cause entry-friendly countries to perpetually grow at a faster rate and the distribution of aggregate productivity to "fan out" in the long-run? While this is an implication of the closed economy version of the model above, the following extension to trade shows that this is not the case. Even though differences in the innovation environment will cause countries to experience differential productivity growth (at the aggregate level) for a while, in the long-run, all countries will grow at the same rate and the cross-country distribution in aggregate TFP will be stable. To see this, consider the following simple extension of the model.

Suppose that the final good is still a Cobb-Douglas aggregate of the continuum of intermediary products, but that different regions specialize in the production of these intermediary products. Specifically, suppose that region r produces a share $s(r)$ of the $[0,1]$-continuum of goods. Letting $y_r(\nu)$ denote the amount of variety ν sourced from region r, the aggregator for the final good can then be written as

$$Y = exp\left(\sum_{r=1}^{R}\int_{\nu\in s(r)} ln\left(y_r(\nu)\right)d\nu\right). \tag{60}$$

While intermediary products are freely tradeable across regions, production factors are immobile. The solution to this model is straight-forward. In particular, let

$$Y_r \equiv \left[exp\left(\int_{\nu\in s(r)} ln\left(y_r(\nu)\right)d\nu\right)\right]^{1/s(r)}$$

be the composite of goods produced in region r and let P_r be the appropriate price index. Using (60) it is easy to see aggregate expenditures across regions will be proportional, i.e.

$$P_r Y_r = s(r)\,Y. \tag{61}$$

It is precisely (61), which will - by affecting region r's terms of trade - create the linkages across regions. Within each region, the economy is exactly the same as the one characterized above. In particular, firms still face demand function with unitary demand elasticity so that they will always be forced to set a limit price. This again implies that the mark-up is the unique state variable for the firms' problem and the distribution of mark-ups determines aggregate allocations. In particular, aggregate (physical) output in region r is given by

$$Y_r(t) = \frac{1}{s(r)} Q_r(t) M_r(t) K_r(t)^{\alpha} L_{P,r}(t)^{1-\alpha}, \tag{62}$$

where $Q_r = \left(exp\left(\int_{\nu \in s(r)} ln\left(q_r(\nu)\right) d\nu\right)\right)^{1/s(r)}$ and $M_r = \frac{exp(-E[\xi_r])}{E[\xi_r^{-1}]}$ as above. Again there exists a unique BGP, where innovation and entry rates are constant but potentially different across regions.

Consider a BGP where the (endogenous) growth rate of aggregate productivity in region r is given by $\dot{Q}_r(t) = Q_r(t) g_r$ and "global" output Y grows at rate g_Y.[45] The equilibrium return to capital in region r is still given by $R_r = \alpha \Lambda_r \frac{s(r)Y}{K_r}$, where $\Lambda_r = E_r\left[\lambda^{-\Delta}\right]$ only depends on the entry intensity in region r. Along the BGP, interest are constant (though not necessarily equal across regions) so that the *regional* capital stock grows at the same rate as *global* output, i.e. $\dot{K}_r(t) = K_r(t) g_Y$. Hence, regional (physical) output grows at rate (see (62)) $g_{Y_r} = g_r + \alpha g_Y$, i.e. productivity growth induces a region-specific component, while capital accumulation is the same across regions. From (60), $Y = \prod_r Y_r^{s(r)}$ so that

$$g_Y = \sum_r s(r) g_{Y_r} = \sum_r s(r) g_r + \alpha g_Y = \frac{1}{1-\alpha} \sum_r s(r) g_r.$$

Hence, the aggregate growth rate is simply a weighted average of the regional rates of productivity growth. (61) then shows that regional output, measured in terms of the numeraire, $P_r Y_r$ grows at rate g_Y so that all regions grow at exactly the same rate despite differences in the growth rates of physical productivity $Q_r(t)$. This is of course due to a deterioration in the terms of trade. In particular, (61) implies that

$$\frac{\dot{P}_r}{P_r} = g_Y - g_{Y_r} = g_Y(1-\alpha) - g_r = \sum_j s(r) g_j - g_r,$$

so that region $r's$ relative prices decline if it grows at a faster rate than the average region. These price effects exactly cancel the heterogeneity in regional productivity growth and cause all regions to grow at the same rate. The mechanism mirrors the one developed in Acemoglu and Ventura (2002). There, countries that accumulate capital faster then the average country see their terms of trade deteriorate. Here, productivity growth plays exactly the same role.

[45] If for example the entry costs χ_r differ across countries, the endogenous productivity growth rate g_r will be different, because the innovation equilibrium is still characterized by the same equilibrium conditions as in the closed economy.

Now suppose that regional product markets in Indonesia were integrated in that way. To see that this would leave the empirical identification intact, note first that the moments regarding entry and the within-firm share of innovators are not affected by the possibility of trading, as labor markets are assumed to clear within each region. Similarly, the growth rate of incumbent firms can be identified from the firm-level data, because $TFPR$ still only depends on the mark-up λ^{Δ} and not on the regional price index. The aggregate growth rate g is different. As I do not use region-specific price deflators in the estimation, the aggregate growth rate will *not* be region-specific but this model with free trade implies that a revenue-based measure of aggregate productivity should grow at the same rate across regions as terms of trade and regional productivity growth are not separately identified. Hence, allowing for free trade would only add an additional cross-region restriction that productivity should grow at the same rate. The heterogeneity in the aggregate growth rate of productivity across regions reported in Table 10 would then be interpreted as pure measurement error. However, the model is still identified in the same way as the closed-economy version above, because even there I restricted the step-size λ to be common across regions.

4 Conclusion

Misallocation of resources across firms reduces economic efficiency and welfare. In an important class of macroeconomic models, misallocation can be measured using firm-level data on productivity. This not only allows to quantify the losses from misallocation but to also test theories. Most theories of misallocation stress the importance of firm-specific barriers to expansion in that some firms are constrained in their input choices. Such constraints can arise from credit-market frictions, differences in non-market access to essential production factors or cronyism that benefits some firms but not others. While these theories differ in their policy implications, they share a common structure of the economic mechanism causing misallocation: some firms are smaller than they ought to be and smaller than they want to be as factors are prevented to flow to these high productivity units. In this paper I proposed a theory of misallocation that is qualitatively different. As firm productivity depends on both prices and quantities, productivity differences might be driven by heterogeneous mark-ups if product markets are not perfectly competitive. High productivity might therefore not be informative about the respective firm being constrained on input markets but rather indicative of market power on the output market. While productive firms are still smaller than they ought to be from a social point of view, they are not smaller than they want to be given their private interests.

If static misallocation is due to firms' market power, the efficiency losses from misallocation depend on the degree of competition which is endogenous. In particular they depend on firms' relative efficiencies as being technologically advanced acts as a shield from competition. Existing firms spend resources on process innovation to increase their mark-up, while entering firms adopt current frontier technologies and thereby keep the mark-up distribution in check. The static degree of misallocation is therefore determined endogenously through the innovation environment. If the equilibrium entry intensity is high (for example because entry barriers like license requirements are low or there is a liquid market for start-up capital), product markets become more competitive, there is little productivity dispersion in the cross-section of firms and the static degree of misallocation is low. If, on the other hand, entrants put little discipline on current producers, there will be pronounced productivity differences across firms and resources will be statically misallocated.

I tested the model's prediction using Indonesian firm-level panel data. I presented reduced form evidence that the extent of market entry by new firms reduces both the level and the dispersion of the cross-sectional productivity distribution, and I show that these results are not driven by regional differences in the quality of the financial system or firm-specific distortions like state ownership. Then I estimated the structural parameters of the model by GMM and showed that the model can account for about 20% of the observed distribution of productivity across producers and that it correctly predicts the regional pattern of productivity dispersion. I also conducted a policy experiment where I decrease entry barriers across regions in Indonesia. A 10% decrease in the costs of entry increases welfare by about 5%. While 10% of these welfare gains are pure static gains in that more intense entry reduces misallocation and hence increases TFP (holding technologies fixed), 30% of the gains are dynamic as improved allocative efficiency induces capital accumulation. The remaining 60% are pure growth effects, as lower entry barriers increase the aggregate growth rate. The empirical regularity that poor countries are characterized by higher cross-sectional productivity dispersion therefore not only suggests that resources are statically misallocated, but might also be a symptom of more fundamental differences in the innovation environment across countries. The aggregate implications of cross-sectional productivity dispersion for cross-country income differences might be much larger than previously appreciated, once such dynamic considerations are taken into account.

5 Appendix

5.1 Data

As explained in the main text, the Statistik Industri dataset contains information on all manufacturing firms in Indonesia with a laborforce of more than 20 employees. For the reduced form analysis of this paper, I only need information on firms' revenues, wagebill, capital stock, location, sector and entry behavior. Additionally I exploit information on the source of firms' investment to measure credit constraints and their ownership structure to identify firms with state ownership. Both revenues and the total wagebill are directly taken from the data. As a measure of capital, I take firms' total assets, which consist of machines, working capital, buildings, vehicles and other forms of fixed capital, all measured at current market prices. The ownership structure is elicited from their capital structure. In particular, each firm reports if the federal or local government or foreign investors own stakes in the firm. The source of finance is also directly observable in the data as firms report the respective share of their investment expenditures, which are financed through foreign direct investment, retained earnings, domestic borrowing, government funds or assets raised on the capital market. Firms also report the share of paid and unpaid workers, which allows me to identify the employment composition. To identify entry, I rely on the panel nature of the data. My measure for the year-province specific entry rate, is the fraction of firms in province r at time t who appear in the data. This measure is not ideal for two reasons. From a conceptual point of view it is not clear why only a firm with at least 20 employees should put competitive pressure on active incumbents. From a measurement point of view, I might label a firm as an entrant if it laid off workers in some years and then started growing again. To address the latter concern, I experimented with other measures, which also condition on the reported age of the firm and other cutoffs and the results were similar. Furthermore, note that this problem only invalidates my empirical strategy if different regions are affected differentially. Regarding the former conceptual concern, La Porta and Shleifer (2009) and McKinsey-Global-Institute (2001) argue that there is very little competition between small, informal firms and members of the formal industrial sector. While these claims mostly concern very small establishments, the 20 employee cutoff might not be too bad a measure of when a firm starts competing within the formal sector. All nominal variables are deflated using the CPI Index from the World Development Indicators. To construct the final sample, some data cleaning steps were necessary. First of all I dropped all establishments, which did not report information on revenues, the wagebill or their asset position. Especially the latter requirement shrinks the sample considerably as roughly a third of plants have missing asset data. To eliminate some concerns about measurement error, I also dropped all firms, which report revenue or employment growth exceeding 200% per year. Finally, following Hsieh and Klenow (2009), I trim the 1% tails of the final laborshare distribution. As most regressions are done on the province-year level, I construct the respective dependent variables using the micro-data. As the manufacturing sector in Indonesia is relatively concentrated, many provinces contain only few firms. Using more firms to construct the averages obviously reduces the noise in the generated variables but also reduces the number of provinces I can use in each year. For the main analysis I drop all province-year cells, which contain less than 25 firms. However, I show below that the results are not dependent on this particular cutoff. In order to have a direct measure of credit constraints for Table 2, I exploit the 1996 survey. The 1996 survey is special in that the Statistik Industri survey was done in conjunction with the economic census. Hence, the 1996 survey contains substantially more information. In particular, firms are asked if they are subject to a major constraint, which they could not overcome and if this constraint refers to the scarcity of capital. They are also asked what efforts they undertook to overcome this capital constraint, e.g. if they applied for a bank loan, sold assets or issued shares. Finally, the structural estimation in Section 3.3 uses the within-firm share of innovation resources to identify the curvature parameter γ. In the data I measure the number of workers employed to generate innovations I as white-collar non-production workers. In the Indonesian data I observe only a crude distinction between production and non-production workers. However, the plant-level data from Chile, which has for example been used by Pavcnik (2002), Moll (2010) and Levinsohn and Petrin (2003), has more detailed occupation characteristics. According to this data, roughly 60% of non-production workers are white-collar. Hence, I measure the firm-level employment share of innovation workers by 0.6 times the share of non-production workers.

The provincial characteristics are constructed using the PODES dataset for the year 1996. The unit of observation is one of the 65.000 villages. For each village, I record the total population, the number of banking branches, an indicator if the village is accessible by asphalted streets, an indicator if the village's main source of income is agriculture and the number of small industrial plants residing in the village. To generate the information on the level of the province, I consider the respective weighted averages of those variables, where the weights are the population sizes of the respective villages.

5.2 Derivation of (20), (22) and (23)

To derive (20), simply note that

$$F_{\Delta}(d;x) = \sum_{i=1}^{d} \mu(i) = x \sum_{i=1}^{d} \left(\frac{1}{1+x}\right)^{i} = x \frac{\frac{1}{1+x}\left(1-\left(\frac{1}{1+x}\right)^{d}\right)}{1-\frac{1}{1+x}} = 1-\left(\frac{1}{1+x}\right)^{d}.$$

As $\Lambda = \left[\int_0^1 \lambda^{-\Delta(\nu)} d\nu\right]$,

$$\Lambda(x) = \sum_{i=1}^{\infty} \lambda^{-i}\mu(i) = \frac{x}{\lambda(x+1)} \sum_{i=0}^{\infty} \left(\frac{1}{\lambda(x+1)}\right)^{i} = \frac{x}{\lambda - 1 + \lambda x},$$

which is (22). Similarly,

$$\int_0^1 \Delta(\nu)\, d\nu = \sum_{i=1}^{\infty} i\mu(i) = x \sum_{i=1}^{\infty} i \left(\frac{1}{1+x}\right)^{i} = \sum_{i=0}^{\infty} \left(\frac{1}{1+x}\right)^{i} = \frac{1+x}{x},$$

so that

$$M(x) = \frac{\lambda^{-\int_0^1 \Delta(\nu) d\nu}}{\int_0^1 \lambda^{-\Delta(\nu)} d\nu} = \frac{1}{\Lambda(x)} \lambda^{-\int_0^1 \Delta(\nu) d\nu} = \lambda^{-\frac{1+x}{x}} \frac{\lambda - 1 + \lambda x}{x},$$

which is (23).

5.3 More general entry process: Derivation of (24)

Suppose that incumbent firms climb one step of the quality ladder with flow rate I. In contrast, entry occurs at a flow rate z but conditional on entry, the new blueprint has a quality advantage of j steps with probability $p(j)$. In that case, the measure of firms having a quality advantage Δ, $\mu(\Delta)$, solves the flow equations

$$\begin{aligned}\dot{\mu}(\Delta,t) &= -(I+z)\mu(\Delta,t)+I\mu(\Delta-1,t)+zp(\Delta) \text{ for } \Delta\geq 2\\ \dot{\mu}(1,t) &= -(I+z)\mu(1,t)+zp(1).\end{aligned}$$

In the stationary distribution we have $\dot{\mu}(\Delta,t)=0$, so that

$$\mu(1)=\frac{zp(1)}{I+z} \text{ and } \mu(\Delta)=\frac{zp(\Delta)+I\mu(\Delta-1)}{I+z}.$$

It can then be verified that the expression for $\mu(\Delta)$ given in (24) solves those equations. Now consider the special case where $p(i)=T\kappa^i$. As $\sum p(i)=1$, we need $T=\frac{1-\kappa}{\kappa}$. For that case, the stationary distribution of productivity gaps Δ (and hence log mark-ups) is given by

$$\mu(\Delta;x,\kappa)=\left(\frac{1}{x+1}\right)^{\Delta}\frac{1-\kappa}{\kappa}\left\{\sum_{i=1}^{\Delta}(\kappa x)^i\left(\sum_{k=0}^{i-1}\binom{i-1}{k}x^{-k}\right)\right\}, \tag{63}$$

where I explicitly note the dependence on the entry intensity x and the "measure of decay" κ. In Figure 5.3 I depict (63) for different values of κ and x. It is seen that a higher level of the entry intensity induces first order stochastic dominance shifts in the distribution for different values of κ.

5.4 Characterization of the BGP

Along the BGP, interest rates are constant and aggregate output grows at a constant rate g_Y. So let us conjecture that both innovator wages $w_I(t)$ and the value functions $V(\Delta,t)$ also grow at rate g_Y and that innovation and entry rates are constant and equal across all sectors, i.e. $z(\Delta,t)=z$ and $I(\Delta,t)=I$. These conjectures will be verified below. To characterize the BGP, conjecture that the value function takes the form of

$$V(\Delta,t)=\kappa(t)-\phi(t)\lambda^{-\Delta}. \tag{64}$$

If V grows at rate g_Y, (72) implies that both $\kappa(t)$ and $\phi(t)$ grow at rate g_Y too. Hence let us normalize all variables by $exp(g_Y t)$ so that they become stationary and t ceases to be a state variables. Using (72) and (9) and

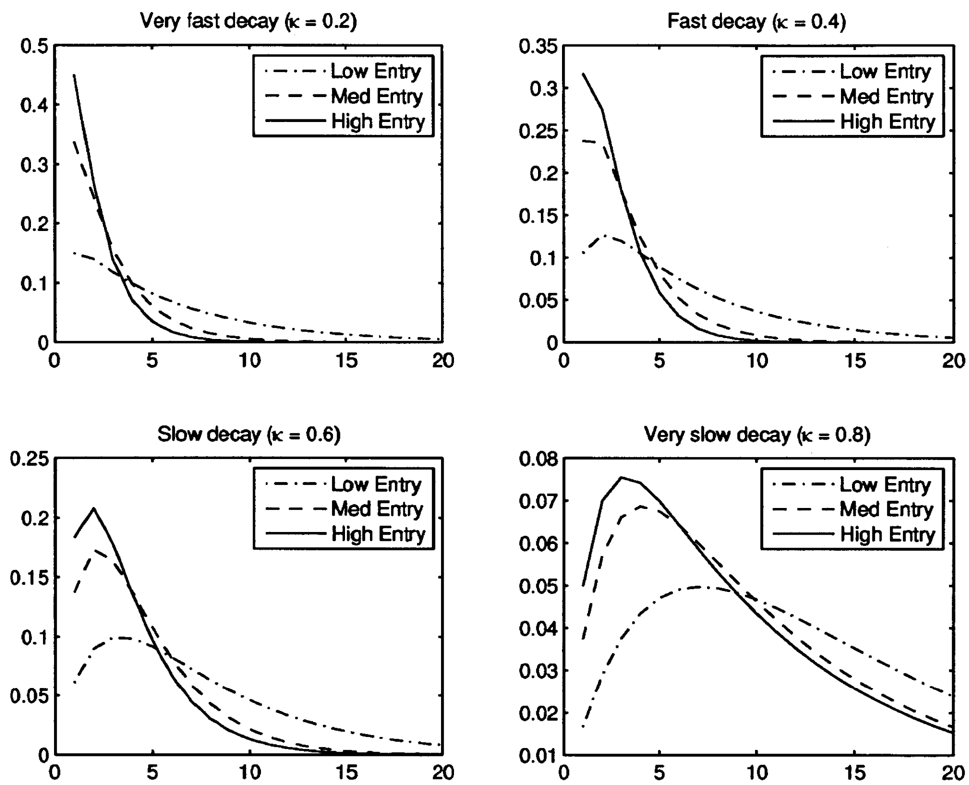
$$V(\Delta+1)-V(\Delta)=-\phi\lambda^{-\Delta-1}+\phi\lambda^{-\Delta}=\phi\lambda^{-\Delta}\frac{\lambda-1}{\lambda},$$

the Bellman equation reads

$$(r+z-g_Y)V(\Delta) = Y\left(1-\lambda^{-\Delta}\right)+\lambda^{-\Delta}\max_I\left\{I\phi\frac{\lambda-1}{\lambda}-w\frac{1}{\varphi}I^{\gamma}\right\}.$$

The optimal innovation flow rate I^* is implicitly defined by the necessary and sufficient FOC

$$\gamma\frac{1}{\varphi}I^{\gamma-1}=\frac{\phi}{w}\frac{\lambda-1}{\lambda}, \tag{65}$$



Notes: The figure shows the implied distribution of mark-ups (63) for different values of the entry intensity x and different values of the decay parameter κ. In particular, I consider $x \in \{0.2, 0.6, 1\}$ and $\kappa \in \{0.2, 0.4, 0.6, 0.8\}$.

Figure 8: Stationary distribution of mark-ups (63) for different (κ, x)

where $\frac{\phi}{w_I}$ is constant along the BGP. Hence, I^* is independent of time and the same for all firms. In particular

$$I\phi\frac{\lambda-1}{\lambda}-w\frac{1}{\varphi}I^{\gamma}=(\gamma-1)\frac{1}{\varphi}I^{\gamma-1}w$$

so that upon substituting (72), we get that

$$\begin{aligned}(r+z-g_Y)\kappa-(r+z-g_Y)\phi\lambda^{-\Delta} &= Y\left(1-\lambda^{-\Delta}\right)+(\gamma-1)w\frac{1}{\varphi}I^{\gamma}\lambda^{-\Delta}.\\ &= Y-\left(Y-(\gamma-1)w\frac{1}{\varphi}I^{\gamma}\right)\lambda^{-\Delta}.\end{aligned}$$

As this equation has to hold for all Δ, we need that

$$\phi=\frac{Y-(\gamma-1)w\frac{1}{\varphi}I^{\gamma}}{r+z-g_Y}. \tag{66}$$

Similarly we get $\kappa=\frac{Y}{r+z-g_Y}$, so that the value function is given by

$$\begin{aligned}V(\Delta,t) &= \kappa(t)-\phi(t)\lambda^{-\Delta}\\ &= \frac{Y(t)}{r+z-g_Y}-\left(\frac{Y(t)-(\gamma-1)w(t)\frac{1}{\varphi}I^{\gamma}}{r+z-g_Y}\right)\lambda^{-\Delta}\\ &= \frac{\left(1-\lambda^{-\Delta}\right)Y(t)+(\gamma-1)w(t)\frac{1}{\varphi}I^{\gamma}\lambda^{-\Delta}}{\rho+z}\\ &\equiv \frac{\pi(\Delta,t)+(\gamma-1)w(t)\Gamma(I,\Delta)}{\rho+z},\end{aligned} \tag{67}$$

where I used that the Euler equation implies that $\frac{\dot{c}(t)}{c(t)}=g_Y=r-\rho$. This verifies (27). For $V(\Delta,t)$ to grow at a constant rate, we need to show that $w(t)$ and output $Y(t)$ grow at a constant rate g_Y. Given the constant innovation rate I and the constant entry rate z, the growth rate g_Y is constant and given by $g_Y=ln(\lambda)(I+z)$. That both w_I and Y grow at this rate, is seen from the free entry condition. We finally have to establish that there exists a solution for I, which is consistent with (73). Using (66), the first-order condition can be rewritten

$$\rho+z \quad = \quad \frac{\eta}{\varphi(1+\eta\chi)}\left(\gamma I^{\gamma-1}(\rho+z)+(\gamma-1)I^{\gamma}\right). \tag{68}$$

As the RHS of (68) is increasing in I and satisfies

$$\begin{aligned}\lim_{I\to 0}\left(\gamma I^{\gamma-1}(\rho+z)+(\gamma-1)I^{\gamma}\right) &= 0\\ \lim_{I\to\infty}\left(\gamma I^{\gamma-1}(\rho+z)+(\gamma-1)I^{\gamma}\right) &= \infty,\end{aligned}$$

there is a unique innovation level I, which is consistent with firm's optimal behavior and the value function $V(\Delta,t)$.

To finally determine the equilibrium innovation level I and the entry intensity x, consider (31) and

(32), i.e.

$$\varphi\frac{1+\eta\chi}{\eta} = \gamma I^{\gamma-1}+\frac{\gamma-1}{\rho+xI}I^{\gamma}\equiv t(x,I) \tag{69}$$

$$\varphi = \frac{x\left(\left(1-\alpha\right)\gamma I^{\gamma-1}\left(\rho+xI\right)\frac{\lambda}{\lambda-1}+\left(\left(1-\alpha\right)\left(\gamma-1\right)+1\right)I^{\gamma}\right)}{\left(\lambda-1\right)+\lambda x}+\varphi\frac{1+\eta\chi}{\eta}xI\equiv h(x,I) \tag{70}$$

From (69),

$$\frac{\partial I}{\partial x}=-\frac{t_x(x,I)}{t_I(x,I)}>0$$

as $t_x(x,I)<0$ and $t_I(x,I)=(\gamma-1)\gamma I^{\gamma-2}+(\gamma-1)\left(\frac{\gamma I^{\gamma-1}(\rho+xI)-I^{\gamma}x}{(\rho+xI)^2}\right)=(\gamma-1)\left[\gamma I^{\gamma-2}+\left(\frac{\gamma I^{\gamma-1}\rho+(\gamma-1)xI^{\gamma}}{(\rho+xI)^2}\right)\right]>0$. From (70),

$$\frac{\partial I}{\partial x}=-\frac{h_x(x,I)}{h_I(x,I)}<0$$

as

$$\begin{aligned}
h_x(x,I) &= \frac{(\lambda-1+\lambda x)\left[\left((1-\alpha)\gamma I^{\gamma-1}(\rho+xI)\frac{\lambda}{\lambda-1}+((1-\alpha)(\gamma-1)+1)I^{\gamma}\right)+x(1-\alpha)\gamma I^{\gamma-1}I\frac{\lambda}{\lambda-1}\right]}{(\lambda-1+\lambda x)^2} \\
&\quad -\frac{x\left((1-\alpha)\gamma I^{\gamma-1}(\rho+xI)\frac{\lambda}{\lambda-1}+((1-\alpha)(\gamma-1)+1)I^{\gamma}\right)\lambda}{(\lambda-1)+\lambda x}+\varphi\frac{1+\eta\chi}{\eta}I \\
&= \frac{(\lambda-1)\left((1-\alpha)\gamma I^{\gamma-1}(\rho+xI)\frac{\lambda}{\lambda-1}+((1-\alpha)(\gamma-1)+1)I^{\gamma}\right)+(\lambda-1+\lambda x)x(1-\alpha)\gamma I^{\gamma-1}I\frac{\lambda}{\lambda-1}}{(\lambda-1+\lambda x)^2} \\
&\quad +\varphi\frac{1+\eta\chi}{\eta}I \\
&> 0 \\
h_I(x,I) &= \frac{x\left[(1-\alpha)\gamma\frac{\lambda}{\lambda-1}\left((\gamma-1)(\rho+xI)I^{\gamma-2}+I^{\gamma-1}x\right)+\gamma((1-\alpha)(\gamma-1)+1)I^{\gamma-1}\right]}{(\lambda-1)+\lambda x} \\
&\quad +\frac{1+\eta\chi}{\eta}x>0.
\end{aligned}$$

Hence, (69) and (70) define two continuous loci with different slopes. Additionally,

$$\lim_{x\to\infty}I^{FE}(x)=I_{\infty}<\infty \text{ and } \lim_{x\to 0}I^{FE}(x)=I_0<I_{\infty}$$

where $I^{FE}(x)$ is the locus implicitly defined by the free entry condition (69).[46] Similarly,

$$\lim_{x\to\infty}I^{LM}(x)=0 \text{ and } \lim_{x\to 0}I^{FE}(x)=\infty,$$

where $I^{LM}(x)$ is locus implicitly defined by the labor market clearing condition (70).[47] Hence, there

[46] In particular, (69) implies that I_{∞} solves $\varphi\frac{1+\eta\chi}{\eta}=\gamma I_{\infty}^{\gamma-1}$ and I_0 solves $\varphi\frac{1+\eta\chi}{\eta}=\gamma I_0^{\gamma-1}+\frac{\gamma-1}{\rho}I_0^{\gamma}$.

[47] To see this, suppose that $\lim_{x\to\infty}I^{LM}(x) > I_{\varepsilon} > 0$. Then,

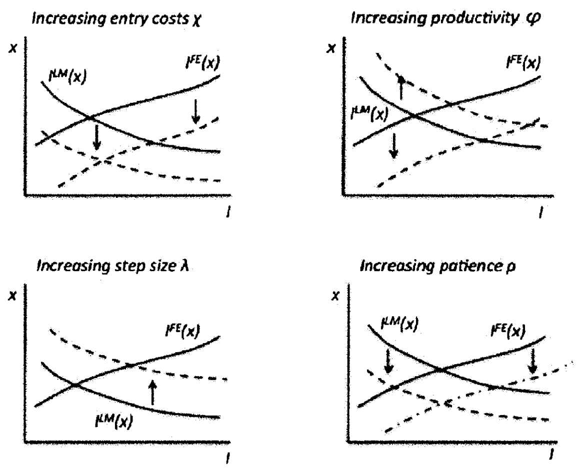


Figure 9: Comparative statics results

is a unique intersection and therefore a unique equilibrium.

To study the comparative statics with respect to the entry costs χ, note that an increase in χ shifts both the I^{FE}-locus and the I^{LM}-locus down. This decreases the equilibrium entry-intensity and has an ambiguous effect on the equilibrium innovation rate. An increase in incumbents' innovation productivity φ shifts the I^{FE}-locus downwards and the I^{LM}-locus upwards and therefore increases the equilibrium innovation rate and has an ambiguous effect on the entry intensity. An increase in the step size λ leaves the I^{FE}-locus unaffected and shifts the I^{LM}-locus upwards. This will increase both the innovation rate and the entry intensity. Finally, an increase in the consumers' discount factor ρ shifts both the I^{FE}-locus and the I^{LM}-locus downwards and hence therefore qualitatively similar effects like an increase in the entry costs χ. Graphically, these results are contained in Figure 9 below.

5.5 Deriving the Budget Constraint (34)

To derive (34), it is useful to think of households having access to two assets to transfer resources across time. They can either save in physical capital or they can save in a mutual fund, who owns the corporate sector of the economy and finances entrants by venture capital. This mutual fund pays

$\lim_{x\to\infty} \frac{x\left((1-\alpha)\gamma I(x)^{\gamma-1}(\rho+xI(x))\frac{\lambda}{\lambda-1}+((1-\alpha)(\gamma-1)+1)I(x)^{\gamma}\right)}{(\lambda-1)+\lambda x} = \infty$ and $\lim_{x\to\infty} \varphi\frac{1+\eta\chi}{\eta}xI(x) = \infty$, which violates (70). Similarly, if $\lim_{x\to 0} I^{LM}(x) < I_{\omega} < \infty$, then $\lim_{x\to 0} \frac{x\left((1-\alpha)\gamma I(x)^{\gamma-1}(\rho+xI(x))\frac{\lambda}{\lambda-1}+((1-\alpha)(\gamma-1)+1)I(x)^{\gamma}\right)}{(\lambda-1)+\lambda x} = 0$ and $\lim_{x\to 0} \varphi\frac{1+\eta\chi}{\eta}xI(x) = 0$.

an interest $r(t)$ for each dollar invested. Letting $A(t)$ be the savings in the mutual fund, the budget constraint of the representative household is given by

$$\dot{A}(t) + \dot{K}(t) + C(t) = w(t) L + (R(t) - \delta) K(t) + r(t) A(t). \tag{71}$$

From (71) it is apparent that no-arbitrage across asset classes directly implies that $R(t) - \delta = r(t)$. Let $\Omega(t)$ be the value of the mutual fund managing the households' savings. As the mutual fund owns all the firms in the economy, hires innovators to fabricate new blue prints and pays an interest rate $r(t)$ to investors, $\Omega(t)$ solves the HJB equation

$$\begin{aligned} r^{MF}(t)\Omega(t) - \dot{\Omega}(t) &= \sum_{i=1}^{\infty} [\pi(i,t) - w(t)\Gamma(i,t)]\mu(i) - w(t) z\left(\frac{1+\chi\eta}{\eta}\right) - r(t) A(t) \\ &\quad + \sum_{i=1}^{\infty} [I(V(i+1,t) - V(i,t)) - zV(i,t)]\mu(i) + zV(1,t), \end{aligned}$$

where $r^{MF}(t)$ is the return of owning the mutual fund. The per-period cash-flows of the mutual fund are given in first row - the mutual fund is the recipient of corporate profits net of innovation expenditures, pays investors an interest rate of $r(t)$ and finances the venture capital for new innovations. The second row contains the capital gains. The first term reflects the changes in the aggregate valuations of the existing firms in the portfolio of the fund. The second term $zV(1,t)$ reflects that the mutual fund owns all new entrants. Now note that

$$\begin{aligned} & \textstyle\sum_{i=1}^{\infty} [I(V(i+1,t) - V(i,t)) - zV(i,t)]\mu(i) + zV(1,t) \\ = & \textstyle\sum_{i=2}^{\infty} V(i,t) I\mu(i-1) - \sum_{i=1}^{\infty} V(i,t)(I+z)\mu(i) + zV(1,t) \\ = & \textstyle\sum_{i=2}^{\infty} V(i,t)[I\mu(i-1) - (z+I)\mu(i)] + V(1,t)[z - (I+z)\mu(1)] = 0, \end{aligned}$$

as μ is stationary so that $I\mu(i-1) - (z+I)\mu(i) = 0$ and $z - (I+z)\mu(1) = 0$ (see (17)). If there is perfect competition among mutual funds, the return of owning a mutual fund has to be zero, i.e. $r^{MF} = 0$. This implies that

$$\begin{aligned} r(t) A(t) &= \sum_{i=1}^{\infty} [\pi(i,t) - w(t)\Gamma(i,t)]\mu(i) - w(t) z\left(\frac{1+\eta\chi}{\eta}\right) \\ &= \sum_{i=1}^{\infty} \pi(i,t)\mu(i) - w(t)\left(z\left(\frac{1+\eta\chi}{\eta}\right) + \sum_{i=1}^{\infty}\Gamma(i,t)\mu(i)\right) \\ &= \Pi(t) - w(t)[L - L_P(t)]. \end{aligned}$$

Substituting this into (71) yields

$$\begin{aligned} \dot{K}(t) + C(t) &= w(t) L + w_I L_S + (R(t) - \delta) K(t) + \Pi(t) - w(t)[L - L_P(t)] \\ &= w(t) L_P + (R(t) - \delta) K(t) + \Pi(t), \end{aligned}$$

which is (34).

5.6 Static Misallocation and Dynamic Incentives

Suppose that firms have access to the same physical productivity q and that the economy has access to an outside technology to produce each variety with productivity q_L. This implies that mark-ups are constant in the cross-section of firms and given by $\xi = \frac{q}{q_L}$. Let us assume for simplicity that $q = \lambda q_L$. Suppose that the innovation technology is given by

$$\Gamma(I,\Delta) = \lambda^{-\Delta}\frac{1}{\varphi}I^{\gamma} + f$$

where f is a fixed cost requirement. I will construct a steady state equilibrium, where neither entry nor innovation occurs. Again, conjecture that the value function is given by

$$V(\Delta) = \kappa - \phi\lambda^{-\Delta}, \tag{72}$$

where V is constant because there is no growth in this economy. The HJB equation now reads

$$r(t)V(\Delta) = \pi(\Delta) + \max_I\left\{I(V(\Delta+1) - V(\Delta)) - w(t)\lambda^{-\Delta}\frac{1}{\varphi}I^{\gamma} - w(t)f\right\},$$

as (by construction) there will not be entry. Again

$$V(\Delta+1) - V(\Delta) = -\phi\lambda^{-\Delta-1} + \phi\lambda^{-\Delta} = \phi\lambda^{-\Delta}\frac{\lambda-1}{\lambda},$$

so that

$$r(t)V(\Delta) = \pi(\Delta) + \lambda^{-\Delta}\max_I\left\{I\phi\frac{\lambda-1}{\lambda} - w(t)\frac{1}{\varphi}I^{\gamma} - w(t)f\lambda^{\Delta}\right\}.$$

The necessary and sufficient FOC for innovation choice is still

$$w(t)\gamma\frac{1}{\varphi}I^{\gamma} = \phi\frac{\lambda-1}{\lambda}I, \tag{73}$$

so that

$$I\phi\tfrac{\lambda-1}{\lambda} - w(t)\tfrac{1}{\varphi}I^{\gamma} - w(t)f\lambda^{\Delta} \quad \leq I\phi\frac{\lambda-1}{\lambda}\frac{\gamma-1}{\gamma} - w(t)f.$$

Now suppose that $I = 0$. Then

$$r(t)V(\Delta) = \pi(\Delta) = \left(1 - \lambda^{-\Delta}\right)Y(t).$$

Along the steady state,

$$Y(t) = Y \text{ and } r(t) = r = \rho$$

so that

$$V(\Delta) = \frac{(1-\lambda^{-\Delta})\,Y}{\rho} = \frac{Y}{\rho} - \lambda^{-\Delta}\frac{Y}{\rho},$$

which shows that

$$\kappa = \phi = \frac{Y}{\rho}.$$

A sufficient condition for this equilibrium is that there is no innovation, i.e.

$$f \;>\; I\frac{\phi}{w}\frac{\lambda-1}{\lambda}\frac{\gamma-1}{\gamma}.$$

The optimal innovation rate solves $I = \left(\frac{\phi}{w}\frac{\lambda-1}{\lambda}\frac{\varphi}{\gamma}\right)^{\frac{1}{\gamma-1}}$, so that

$$I\frac{\phi}{w}\frac{\lambda-1}{\lambda}\frac{\gamma-1}{\gamma} \;=\; \left(\frac{Y}{w}\frac{1}{\rho}\frac{\lambda-1}{\lambda}\frac{1}{\gamma}\right)^{\frac{\gamma}{\gamma-1}}(\gamma-1)\,\varphi^{\frac{1}{\gamma-1}}.$$

In a steady state equilibrium, the entire laborforce is employed in the production sector, so that

$$w = (1-\alpha)\,Y\Lambda.$$

Hence, a necesary condition for firms to not be willing to spend resources on innovation is that

$$f > \left(\frac{\lambda-1}{1-\alpha}\frac{1}{\rho}\frac{1}{\gamma}\right)^{\frac{\gamma}{\gamma-1}}(\gamma-1)\,\varphi^{\frac{1}{\gamma-1}}.$$

This is a parametric condition, which is satisfied if the fixed costs f are high, the productivity φ is low, the innovation step-size λ is low and the discount rate ρ is high, so that innovations, which pay off in the future, are valued little. For this equilibrium to occur, we also require that there are no entry incentives. There is no entry if

$$w \geq \eta\,(V(1) - \chi w) = \frac{\eta}{1+\chi\eta}V(1) = \frac{\eta}{1+\chi\eta}\frac{\lambda-1}{\lambda}\frac{Y}{\rho},$$

Rearraning terms yields

$$\frac{1+\chi\eta}{\eta} \;\geq\; \frac{\lambda-1}{\lambda}\frac{1}{\rho}\frac{Y}{w} = \frac{1}{\rho}\frac{\lambda-1}{1-\alpha}.$$

Again, this is a parametric condition which is satisfied if the discount rate is high, the innovation step-size is low, the entry costs χ are high and the entry productivity η is low. In such a world there is no static misallocation, as mark-ups and hence measured productivity are equal in the cross-section of firms. However, the economy is not particularly efficient, because there is no growth.

5.7 A Simple Model of Credit Constraints

Given the environment described in the text, firms solve the problem

$$\begin{array}{rc} & \max_{(k,l)} Y(t)\left(ql^{1-\alpha}k^{\alpha}\right)^{\frac{\sigma-1}{\sigma}} - wl - Rk \\ \text{s.t.} & wl + Rk \leq \theta A. \end{array}$$

Letting μ be the multiplier on the collateral constraint, the two optimality conditions are given by

$$\begin{aligned} Y\frac{\sigma-1}{\sigma}(1-\alpha)\left(ql^{1-\alpha}k^{\alpha}\right)^{\frac{\sigma-1}{\sigma}}\frac{1}{l} &= (1+\mu)w \\ \frac{k}{l} &= \frac{\alpha}{1-\alpha}\frac{w}{R}. \end{aligned}$$

If $\mu > 0$, then $\theta A = wL + Rk = \frac{w}{1-\alpha}L$ so that

$$l^{C}(a,q) = \frac{1-\alpha}{w}\theta A$$

If $\mu = 0$, then

$$l^{U}(q) = \frac{1-\alpha}{w}\left(Y\frac{\sigma-1}{\sigma}\right)^{\sigma}\left(\frac{1}{\psi(w,R)}\right)^{\sigma-1}q^{\sigma-1}.$$

where $\psi(w,R) = \left(\frac{w}{1-\alpha}\right)^{1-\alpha}\left(\frac{R}{\alpha}\right)^{\alpha}$. The solution to the problem is therefore given by

$$\begin{aligned} l(a,q) &= \min\left\{l^{C}(a), l^{U}(q)\right\} \\ &= \frac{1-\alpha}{w}\min\left\{\theta A, \left(\frac{\sigma-1}{\sigma}\right)^{\sigma}Y^{\sigma}\left(\frac{1}{\psi(w,R)}\right)^{\sigma-1}q^{\sigma-1}\right\} \end{aligned}$$

Hence, the firm is constrained if

$$\tilde{A} = \frac{A}{q^{\sigma-1}} < \frac{1}{\theta}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma}Y^{\sigma}\left(\frac{1}{\psi(w,R)}\right)^{\sigma-1} \equiv \overline{A}(q,w,R,Y,\theta),$$

which is (40). Productivity is given by $TFPR = \frac{py}{l^{\alpha}k^{1-\alpha}} = pq$, where $p(\nu) = Y^{\frac{1}{\sigma}}y^{-\frac{1}{\sigma}}$. If firms are unconstrained,

$$y^{U}(\nu) = qk^{U}(q)^{\alpha}l^{U}(q)^{1-\alpha} = \left(Y\frac{\sigma-1}{\sigma}\right)^{\sigma}\left(\frac{1}{\psi(w,R)}\right)^{\sigma}q^{\sigma},$$

so that $p^{U}(q) = \left(\frac{\sigma}{\sigma-1}\right)\frac{\psi(w,R)}{q}\left(\frac{1}{Y}\right)^{\frac{\sigma-1}{\sigma}}$. This implies that

$$TFPR^{U} = \frac{\sigma}{\sigma-1}\frac{\psi(w,R)}{Y^{\frac{\sigma-1}{\sigma}}},$$

i.e. TFPR is constant. If firms are constrained, $y^C = \frac{1}{\psi(w,R)} q\theta a$, so that $p^C(a,q) = (Y(t)\psi(w,R))^{\frac{1}{\sigma}} \left(\frac{1}{q\theta a}\right)^{\frac{1}{\sigma}}$. Hence,

$$TFPR^C(a,q) = (Y(t)\psi(w,R))^{\frac{1}{\sigma}} \left(\frac{1}{\theta}\right)^{\frac{1}{\sigma}} \left(\frac{q^{\sigma-1}}{a}\right)^{\frac{1}{\sigma}}.$$

To get (41), note that

$$\begin{aligned} ln(TFPR(A,q)) &= \left(1 - 1\left[\tilde{A} < \overline{A}(q,w,R,Y,\theta)\right]\right) ln\left(TFPR^U\right) + 1\left[\tilde{A} < \overline{A}(q,w,R,Y,\theta)\right] ln\left(TFPR^C(A,q)\right) \\ &= ln\left(TFPR^U\right) + 1\left[\tilde{A} < \overline{A}(q,w,R,Y,\theta)\right] \frac{1}{\sigma}\left[ln\left(\frac{q^{\sigma-1}}{A}\right) - ln\left(\frac{1}{\overline{A}(q,w,R,Y,\theta)}\right)\right] \\ &= \text{const} + \frac{1}{\sigma} 1\left[\tilde{A} < \overline{A}(q,w,R,Y,\theta)\right]\left[ln\left(\frac{\overline{A}(q,w,R,Y,\theta)}{\tilde{A}}\right)\right], \end{aligned}$$

as by construction $TFPR^C\left(\overline{A}(q,w,R,Y,\theta)\, q^{\sigma-1}, q\right) = TFPR^U$.

Chapter 2

Labor Migration and Biased Technological Progress: Evidence from Germany's Post-War Population Transfer[1]

1 Introduction

How do firms respond to large changes in labor supply? In the short-run, falling wages will induce firms to substitute towards the abundant factor. In the long-run however, firms' labor demand will depend on their technological adoption decisions. If firms' technological choices are affected by the labor supply they face, labor supply shifts will induce movements in aggregate labor demand. The long-run consequences of supply-shocks can therefore be very different from their short-term implications once technological adoption has been taken into account. It is this reasoning that is at the heart of the literature on endogenous technological bias (Acemoglu, 2007).[2]

In this paper, I exploit historical data on large supply shocks to test for the existence of endogenous bias. In particular, I study the long-term effects of the post-World War II expulsion of ethnic Germans from Eastern Europe and their subsequent transfer to Western Germany. This historical episode is promising to test for the existence of biased technological adoption. First of all, the induced changes in labor supply are of a very large magnitude. Between 1945 and 1947, more than 8 million people were expelled. This amounted to 20% of the population in Western Germany. Hence, the market size for potential new technologies which are biased towards refugee labor is large. Secondly, the allocation of refugees across counties was unrelated to refugees' skill set. This is clearly important to attribute technological adoption decisions to the initial supply shifts. Third, there is large variation in the number of refugees allocated to different local labor markets. And finally, the initial increase in labor supply induced by the

[1]I am especially grateful to Daron Acemoglu and Kevin Murphy for their precious advice. I also thank David Autor, Abhijit Banerjee, Joaquin Blaum, Davide Cantoni, Esther Duflo and Richard Hornbeck for helpful discussions and seminar participants at MIT and Harvard University for their suggestions. Furthermore I thank Thomas Grosser for his help with the historical data sources, Christian Drews for support with IABS data, Paul Luettinger for his hospitality during my stay at the GESIS in Mannheim and Christina Gathman and Uta Schoenberg for providing me with the crosswalk to match the BBIB and IABS data on the occupation level.

[2]An innovation is biased towards a production factor, if it increases its marginal product relative to other factors.

inflow of refugees had a long persistence as migration within Germany was severely restricted by the respective Military Governments until the early 1950s. That labor supply shocks were long-lasting is important to find evidence of technological adoption, which arguably requires time of adjustment.

To guide the empirical analysis, I first develop a simple theoretical framework that captures this historical episode in a reduced form way. In particular a study a static two-sector Roy-economy with labor being the single factor of production. The theory has two crucial ingredients. To parametrize the sensitivity of technological adoption, I model one sector as having aggregate increasing returns due to a demand externality stemming from monopolistic competition (Krugman, 1980; Matsuyama, 1995). The degree of increasing returns determines the response of entry of new varieties with respect to changes in sector-specific labor supply and hence parametrizes both the elasticity of labor demand and the degree of technological bias. On the supply side, I assume that refugees and natives are heterogeneous in their sector-specific skills, i.e. differ in comparative advantage. Differences in the share of refugees in the labor force will therefore induce shifts in aggregate labor supply, which is not homogeneous across sectors.

The model has three main implications. Due to differences in comparative advantage, natives and refugees will sort into different occupations. As refugees enter the labor market in Western Germany with a mismatched skill set, refugee-rich counties are predicted to have higher employment shares in industries, where job-specific experience and human capital are of little importance. These shifts in labor supply trigger technology adoption, which is modelled as the entry of new varieties. The strength of this technological response is governed by the degree of increasing returns. If these increasing returns are high enough, there will be strong equilibrium bias in that the long-run demand function for skills is *upward* sloping. If that is the case, the model predicts that refugee-rich counties should have especially high wages in refugee-intensive industries. Finally, the model makes predictions about the cross-sectional difference in earnings between refugees and natives. As refugees have a comparative advantage in unskill-intensive industries but no absolute advantage in any industry, equilibrium sorting implies that refugees will earn lower wages. In particular, these lower wages stem purely from differences in occupational choice (i.e. refugees end up working in low-paying occupations) and not from differences in individual-specific efficiency.

It is these implication, which I will take to the data. Using various sources of microdata from the early 60s to the mid 70s, the empirical analysis proceeds in two steps. I first consider the cross-sectional evidence on differences in occupational sorting and earnings between natives and refugees. I find that the expulsion caused substantial sectoral and occupational change for the population of refugees. While the pre-war employment share in manufacturing was only 30%, more than 50% of the refugee population worked in manufacturing industries after the expulsion. I also find that this sectoral shift is accompanied by occupational downgrading in that the share of workers being employed in unskilled occupations increased by fifteen percentage points between its pre-war level in 1939 and 1950. Strikingly, this adjustment is only present for the population of refugees, i.e. neither the sectoral and occupational patterns show any difference between 1939 and 1950 for the native population. Through the lens of the theory, I interpret this finding as refugees having a comparative advantage in unskilled occupations in the manufacturing sector and as the elasticity of labor demand in these industries to be relatively high. While these changes are patterns in the cross-section of workers, my data allows me to decompose them into changes across cohorts and changes within the employment life-cycle. I show that essentially all of the adjustment takes place within the life-cycle of individual workers and that for the vast majority of them persists for the remainder of their working life. Turning to the wage data, I find long-lasting negative effects of the initial displacement on individual earnings. 15 years after the initial expulsion, refugees' wages are still 10% lower and 70% of this monetary penalty is explained through the differential sorting across sector and occupation cells. Additionally, in the cross-section, this adverse effect is less pronounced for young refugees who began their career in Western Germany after the expulsion and show less differences in their sorting across sector-occupation cells compared to the native population. Finally, in 1971, the average wage-penalty for refugees dropped to merely 5%, reflecting the fact that older cohorts, for which comparative disadvantage was worst, left the labor force. Hence, differences in occupational choice account for a large share of the differences in earnings between refugees and natives.

Then I turn to the level of local labor markets. Using wage data from social security records in 1975, I am able to match individual wages to local labor markets and to the initial allocation of refugees in 1950. I augment this data with information on detailed job characteristics, which allow me to measure the importance of formal human capital, on-the-job training and the reliance on unskill-biased technologies at the job-level. This allows me to study the reduced form effects of refugee inflows on wages and employment patterns at the level of local labor market as a whole and for specific occupational cells. Both the OLS and an instrumental variable strategy shows that the initial refugee allocation in 1950 is strongly positively correlated with the size of the manufacturing sector and the employment shares of unskill-intensive occupations in 1975. I also find that refugee inflows are robustly positively correlated with average wages and that this effect is heterogeneous across the population. In particular, young and unskilled workers benefit more from being in a refugee-rich county and refugee rich counties have especially high wages in jobs, which rely on on-the-job training and use unskill-biased technologies. These results are evidence of the existence of strong equilibrium bias in that the long-run relative demand schedule is upward sloping. Through the lens of the theory, they imply that both the elasticity of substitution across industries and the degree of increasing returns are high.

Related Literature This paper is related to three strands of the literature. Most importantly, the paper is related to the work on endogenous innovation and in particular biased technological progress (Acemoglu, 2007). According to this theory, increases in the supply of specific factors of production induce innovation, which is biased towards that factor. This framework has been used to study the evolution of inequality across skill groups (Acemoglu, 1998, 2003), the short-run effects of unskilled migration (Lewis, 2011) and the innovation incentives of producers after adverse shocks to input supply (Hanlon, 2011). My paper complements these studies by showing that the initial expulsion "pushed" refugees predominantly into unskilled occupations and that refugee-rich counties have higher wages in precisely those occupations two decades later. Also, on the theory side, I construct a simple model that features endogenous bias towards different sectors (not skills). As labor supply is not exogenously differentiated by sector, the equilibrium degree of bias and the sorting of workers across occupations is jointly determined.

Secondly, there is an extensive literature on migration, which is for example reviewed in Borjas (1999). The majority of papers in that literature are interested in estimating the consequences of migration on the outcomes of natives, in particular the short-run effects - see for example Borjas (2003), Card (1990) or especially Friedberg (1997), who studies the effects of mass migration into the Israeli labor market. The focus of this paper is different in that I am exclusively interested in the long-run and that the data I am using allows me to study employment histories over a span of three decades.

Finally, the paper also contributes to a large literature in development economics on labor mobility across sectors. Since the famous model of Lewis (1954), many papers address potential frictions or inefficiencies in mobility across sectors of production and the welfare consequences thereof (Banerjee and Newman, 1998; Acemoglu and Zilibotti, 1999; Lagakos and Waugh, 2011). My theory stresses an externality similar to Matsuyama (1992), where in the equilibrium too few workers are allocated to the manufacturing sector. However, technological progress in my model is not exogenous. In contrast, the sectoral allocation of workers determines the equilibrium market size and therefore firms' investment incentives (Murphy, Shleifer, and Vishny, 1989; Matsuyama, 1995; Krugman, 1991). This link between changes in labor supply and endogenous technological adoption is absent from many other papers regarding episodes of structural change shifting workers across different sectors of production (Caselli and Coleman II, 2001).

The rest of the paper is structured as follows. In the next section I present a simple theoretical framework to guide the empirical analysis, which is contained in section three. I first describe the historical background of the expulsion. In particular, I discuss the way the initial allocation of refugees across counties was decided upon and argue that internal migration was limited so that the initial allocation of refugees persisted for a long time. Then I study the post-expulsion outcomes. I first present the evidence at the individual level, i.e. the differences between natives and refugees in the 1960s and 1970s. Then I turn to the long-run consequences for local labor markets and test for the existence of endogenous bias. Section four offers some concluding remarks.

2 The Model

In this section I present a simple framework to study the long-run effects of refugee inflows on local labor markets. The model is a two-sector economy with labor being the single factor of production. Hence, labor supply is not exogenously differentiated by sector-specific skills, but the sorting across sectors of production is determined in equilibrium as in Roy (1951). The model has two crucial ingredients. On the labor demand side, I allow for technological adoption, which I for simplicity model as an aggregate demand externality as in Krugman (1980) and Murphy, Shleifer, and Vishny (1989). This structure delivers a very tractable model of endogenous technological adoption, which is (potentially) biased towards different factors of production. On the labor supply side, I assume that individuals, i.e. refugees and natives, differ in their comparative advantage across sectors. This implies that refugee inflows induce shifts in aggregate labor supply, which are not a pure scale effect, but heterogeneous across sectors.

Formally, I consider a static economy, which is populated by L^N native individuals and L^R refugees. The only source of heterogeneity is their sector-specific talent. In particular, agents have homogeneous preferences, which are given by

$$U = \left(\gamma c_S^{\frac{\varepsilon-1}{\varepsilon}} + (1-\gamma)\, c_F^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{1}$$

where c_S and c_F denote the respective consumption levels of the two goods and ε is the elasticity of substitution. For lack of a better term, I will refer to c_S as the "skilled" good and c_F as the "factory" good. In particular, I think of the factory good as requiring little specific experience and hence as the sector where refugees have a comparative advantage. The two sectors differ in their production technology. In particular, the specific good is produced using a constant returns technology, i.e.

$$Y_S = AE_S,$$

where E_S is the number of efficiency units employed in production and A parametrizes the productivity. Factory production is subject to increasing returns as it requires a fixed setup cost. I model the factory commodity as a standard CES composite of differentiated varieties, which are provided by monopolistically competitive firms. If there is a measure N of varieties, total production of the factory good is given by

$$Y_F = N^{\theta - \frac{1}{\sigma-1}} \left(\int_{\omega=0}^{N} y(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}, \tag{2}$$

where $y(\omega)$ is the quantity of intermediate good ω. As in Acemoglu, Antras, and Helpman (2007), the parameter θ controls the degree of increasing returns. In particular, suppose that $y(\omega) = y$ so that $Y_F = yN^{\theta - \frac{1}{\sigma-1}} N^{\frac{\sigma}{\sigma-1}} = yN^{1+\theta}$. For $\theta = 0$, there are no increasing returns in the aggregate as love-for-variety effects are absent. For $\theta = \frac{1}{\sigma-1}$, this expression reduces to the standard "CES-monopolistic-competition"-framework. The degree of increasing returns θ will turn out to crucially determine the shape of the aggregate demand function. Hence, it is useful to be able to separate the degree of increasing returns θ from the demand elasticity σ, which determines firms' monopoly power.

The production of intermediate products requires an initial setup cost of κ units of the factory good and is then given by a CRS production function

$$y(\omega) = \varphi l(\omega), \tag{3}$$

where $l(\omega)$ is the labor input of firm ω (measured in efficiency units) and φ is a productivity parameter. As I am interested in the long-term properties of the economy, I assume that there is a large number of potential entrants, so that profits of intermediate producers will be driven to zero.

On the labor supply side, I assume that refugees and natives differ in their sector specific skills. Let $e^i = \left(e^i_S, e^i_F\right)$ be the number of efficiency units, individual i can provide to sector S and F respectively. As I think of the factory sector as requiring little specific knowledge, I assume that refugees have a *comparative* advantage in factory production and a (weak) absolute disadvantage in both sectors. In particular, I assume that

$$e^N = (e_S, e_F) \text{ and } e^R = (\chi e_S, e_F), \tag{4}$$

where $\chi < 1$. That both natives and refugees are equally productivity when employed in a factory, is conceptually useful because it allows me to focus on firms' responses to occupational sorting and abstract from pure efficiency effect. An equilibrium in this economy has the usual definition.

Definition 5. Consider the economy above. An equilibrium is a set of prices and wages $[P_S, [p(\omega)]_\omega, w_S, w_F]$, a measure of factory firms N, labor demands $[E_S, [l(\omega)]_\omega]$ and consumption demands $[c_S, [c_F(\omega)]_\omega]$ such that firms maximize profits, consumers maximize utility, markets clear and the allocation is consistent with free entry.

To characterize the equilibrium, it is useful to focus on the labor market. Hence, consider first aggregate labor demand. Given the CES structure of the factory composite, intermediary producers will set a constant mark-up, i.e. intermediary prices are given by $p(\omega) = \frac{\sigma}{\sigma-1}\frac{1}{\varphi}w_F$, where w_F is the wage per efficiency unit in the factory sector. Together with (2) this implies that the price-index for the factory composite is given by

$$P_F \equiv \frac{\left(\int p(\omega)^{1-\sigma} d\omega\right)^{\frac{1}{1-\sigma}}}{N^{\theta-\frac{1}{\sigma-1}}} = \frac{\sigma}{\sigma-1}\frac{1}{\varphi}w_F N^{-\theta}. \tag{5}$$

Total profits of each intermediary producers are $\pi(\omega) = \frac{1}{\sigma-1}\frac{w_F}{\varphi}y(\omega) - P_F\kappa$, so that free entry requires that

$$y(\omega) = \varphi(\sigma-1)\kappa\frac{P_F}{w_F} = \sigma\kappa N^{-\theta}, \tag{6}$$

where the last equality uses (5). Net production in the factory sector (after taking account of the fixed cost expenses) is given by

$$X_F = Y_F - \kappa N = yN^{1+\theta} - \kappa N = (\sigma-1)\kappa N. \tag{7}$$

Letting E_F be the total number of efficiency units allocated to the "assembly lines" of this economy, we have

$$E_F = \int_\omega^N \frac{y(\omega)}{\varphi}d\omega = \frac{y}{\varphi}N = \frac{\sigma\kappa}{\varphi}N^{1-\theta}. \tag{8}$$

This expression is noteworthy, because it governs the degree of technological adoption in this economy. In particular, the degree of entry is increasing in the total factory labor supply and the sensitivity of entry with respect to the "market size" E_F is governed by the degree of increasing returns. It is this reduced form formulation of technological adoption, which determines the shape of aggregate labor demand as (7) and (8) imply that $X_F \propto E_F^{\frac{1}{1-\theta}}$. To solve for the labor demand function, note that the demand system from (1) implies that

$$\left(\frac{c_S}{c_F}\right)^{-\frac{1}{\varepsilon}} = \frac{1-\gamma}{\gamma}\frac{P_S}{P_F} = \frac{1-\gamma}{\gamma}\frac{\sigma-1}{\sigma}\frac{\varphi}{A}\frac{w_S}{w_F}N^{\theta},$$

where I used the fact that $P_S = \frac{1}{A}w_S$. Rearranging terms yields

$$\frac{w_F}{w_S} = \frac{1-\gamma}{\gamma}\frac{\sigma-1}{\sigma}\frac{\varphi}{A}N^{\theta}\left(\frac{c_F}{c_S}\right)^{-\frac{1}{\varepsilon}}. \tag{9}$$

This expression shows that N is factory biased in that N increases the relative wage of factory workers holding quantities fixed. Note also that according to (8), N is increasing in E_F. Hence, this economy features weak equilibrium bias as long $\theta > 0$ (Acemoglu, 2007). I.e. whenever there are aggregate increasing returns in factory production, an increase in labor supply induces entry, which is biased towards employees in the factory sector. To derive the aggregate labor demand function, note that good market clearing requires that

$$\frac{c_F}{c_S} = \frac{X_F}{Y_S} = \frac{\kappa(\sigma-1)N}{AE_S}.$$

Substituting this in (9) and using the relationship between N and E_F yields the aggregate labor demand function

$$\begin{aligned}\frac{w_F}{w_S} &= \frac{1-\gamma}{\gamma}\frac{\sigma-1}{\sigma}\frac{\varphi}{A}N^{\theta}\left(\frac{\kappa(\sigma-1)N}{AE_S}\right)^{-\frac{1}{\varepsilon}}. \\ &= \Phi E_S^{\frac{(\varepsilon-1)\theta}{\varepsilon(1-\theta)}}\left(\frac{E_F}{E_S}\right)^{\frac{\theta\varepsilon-1}{\varepsilon(1-\theta)}},\end{aligned} \tag{10}$$

where $\Phi = \frac{1-\gamma}{\gamma}\left(\frac{\sigma-1}{\sigma}\frac{\varphi}{A}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{\varphi}{\kappa\sigma}\right)^{\frac{(\varepsilon-1)\theta}{\varepsilon(1-\theta)}}$ is a constant. There are two crucial properties of labor demand in this economy. First of all, note that

$$\frac{\partial ln\left(\frac{w_F}{w_S}\right)}{\partial ln\left(E_S\right)}\Big|_{\frac{E_F}{E_S}} = \frac{(\varepsilon-1)\theta}{\varepsilon(1-\theta)},$$

which shows that (holding $\frac{E_F}{E_S}$ fixed) an increase in market size is factory-biased whenever the economy has increasing returns in the aggregate. This is intuitive given the existence of weak equilibrium bias encapsulated in (8) and (9). Additionally,

$$\frac{\partial ln\left(\frac{w_F}{w_S}\right)}{\partial ln\left(\frac{E_F}{E_S}\right)}|_{E_S} = \frac{\theta\varepsilon - 1}{\varepsilon(1-\theta)},$$

which is the slope of relative labor demand, holding the size of the skilled industry constant. Given E_S, the aggregate relative demand function is *upward* sloping as long as

$$\varepsilon\theta > 1,$$

i.e. as long as the degree of increasing returns is sufficiently high and the relative price effect sufficiently weak, which is the case if the elasticity of substitution is high. This property is a form of strong equilibrium bias, where an increase in factory employment induces firms to enter, which in turn increases labor demand and hence the equilibrium wage. It is also clearly seen from (10) that *both* increasing returns and a relatively high elasticity of substitution are responsible for this result. If $\theta = 0$, i.e. there are no aggregate increasing returns, labor demand reduces to

$$\frac{w_F}{w_S} = \frac{1-\gamma}{\gamma}\left(\frac{\sigma-1}{\sigma}\frac{\varphi}{A}\right)^{\frac{\varepsilon-1}{\varepsilon}}\left(\frac{L_F}{L_S}\right)^{\frac{-1}{\varepsilon}},$$

which is just the usual CES demand system with constant returns to scale production. Similarly, in the Cobb-Douglas case of $\varepsilon = 1$ we get that

$$\frac{w_F}{w_S} = \frac{1-\gamma}{\gamma}\left(\frac{L_F}{L_S}\right)^{-1},$$

so that labor demand does not depend on the degree of aggregate returns θ.

To characterize the equilibrium and derive the empirical implications, we of course need the aggregate labor supply schedule in this economy. With sector-specific efficiencies given by (4), it follows directly that equilibrium wages have to satisfy

$$1 \leq \frac{w_S e_S}{w_F e_F} \leq \frac{1}{\chi}$$

If $\frac{w_S e_S}{w_F e_F} < 1$, even natives prefer to work in the factory sector and if $\frac{w_S e_S}{w_F e_F} > \frac{1}{\chi}$, refugees supply their labor to skilled producers despite their relatively low efficiency. Letting L_j^i be the labor supply of type $i = N, R$ in sector $j = S, F$, labor market clearing requires that

$$\begin{aligned} E_F &= e_F L_F^R + e_F L_F^N \\ E_S &= \chi e_S L_S^R + e_S L_S^N, \end{aligned}$$

with $L^j = L^j_S + L^j_F$. The interesting equilibria in this economy are characterized by $1 < \frac{w_A e_A}{w_M e_M} < \frac{1}{\chi}$, as otherwise relative wages are fully determined from the labor supply side. As long as parameters are such that this condition holds, both natives and refugees strictly prefer to work in their respective sector of comparative advantage. Normalizing $e_S = e_F = 1$, labor supply is given by

$$L^R_S = L^N_F = 0 \text{ and } E_F = L^R \text{ and } E_S = L^N, \tag{11}$$

and equilibrium wages follow from (10) and (11) as

$$\frac{w_F}{w_S} = \Phi \left(L^N\right)^{\frac{(\varepsilon-1)\theta}{\varepsilon(1-\theta)}} \left(\frac{L^R}{L^N}\right)^{\frac{\theta\varepsilon-1}{\varepsilon(1-\theta)}}. \tag{12}$$

To solve for the level of wages, let us take the final consumption basket as numeraire, i.e. let

$$P = 1 = \left[\gamma^\varepsilon P_S^{1-\varepsilon} + (1-\gamma)^\varepsilon P_F^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}.$$

Using (12), $P_S = \frac{1}{A} w_S$ and P_F given in (5), wages are given by

$$w_S = A\gamma^{\frac{\varepsilon}{\varepsilon-1}} \left[1 + \Phi \left(\left(L^N\right)^{\frac{\theta}{1-\theta}} \left(\frac{L^R}{L^N}\right)^{\frac{1}{1-\theta}}\right)^{\frac{\varepsilon-1}{\varepsilon}}\right]^{\frac{1}{\varepsilon-1}} \tag{13}$$

$$w_F = \Phi \left(L^N\right)^{\frac{(\varepsilon-1)\theta}{\varepsilon(1-\theta)}} \left(\frac{L^R}{L^N}\right)^{\frac{\theta\varepsilon-1}{\varepsilon(1-\theta)}} w_S. \tag{14}$$

The main properties of the equilibrium are contained in the following proposition.

Proposition 6. *Consider the economy above and suppose that $1 < \frac{w_S}{w_F} < \frac{1}{\chi}$. Then, for a given level of the native population L^N, a higher refugee share $\frac{L^R}{L^N}$*

1. *Increases the share of employment in the factory sector $\frac{E_F}{E_S} = \frac{L^R}{L^N}$*
2. *Increases the relative factory wage $\frac{w_F}{w_S}$ if and only if $\varepsilon\theta > 1$*
3. *Increases the level of skilled wages w_S*
4. *Increases the level of factory wages w_F if $\varepsilon\theta > 1$.*

Furthermore, refugees earn less than natives.

Proof. 1.-4. follow immediately from (11),(12),(13) and (14). That refugees earn lower wages follows from the fact that equilibrium wages satisfy $\frac{w_S}{w_F} > 1$. □

Proposition 6 contains both the main properties of the equilibrium and will guide the empirical work in the next section. The mechanism for these results is the interaction between biased technological adoption on the demand side and sorting according to comparative advantage on the supply side. With refugees having a comparative advantage in the production of the factory good, it is not surprising that they will (in equilibrium) work in that sector and that a larger share of refugees in the population will increase wages for skilled workers. That a large market for factory employment might however especially benefit factory workers stems from the technological bias induced by firm entry. If both the degree of increasing returns and the elasticity of substitution are high enough, the labor demand curve is upward sloping and wages in the factory sector might be higher, the more refugees are present in the local labor market. This of course also implies that the level of wages for factory workers is increasing in the share of refugees. Note that the "strong bias" condition $\varepsilon\theta > 1$ is sufficient but not necessary for factory wages to increase in the refugee share. Finally, if we look at the cross-section within a labor market, refugees earn always less than natives, because of their lack of opportunities. While natives can always take advantage of wage differentials across sectors, refugees lack the ability to be profitably employed in non-factory production. In equilibrium, they will therefore earn lower wages than their native peers. This cross-sectional effect stems purely from the assumption of occupational choice. If skills were exogenously differentiated by sector, the relative wage across sector would of course depend on the full set of exogenous parameters of the model, in particular the respective productivity levels.

Note also that Proposition (6) has an interesting corollary.

Corollary 7. *If $\varepsilon\theta > 1$, welfare is increasing in the share of refugees holding the level of the native population fixed.*

This corollary follows directly from the fact that $\varepsilon\theta > 1$ is sufficient for wages in both sectors to increase. The reason why refugee inflows can increases aggregate welfare in this economy, is again linked to the aggregate increasing returns. The equilibrium allocation in this economy is not efficient, because firms do not internalize the aggregate returns of expanding varieties when making their entry choices. Hence, in equilibrium too much labor will be allocated towards the skilled sector and a benevolent government would want to tax labor in the skilled industry to induce workers to switch occupations. Refugees with their comparative disadvantage in the sector, which is inefficiently large, perform exactly this role - in fact, their comparative disadvantage is equivalent to a tax from their point of view. Having more refugees in the local economy might therefore increase welfare by improving allocative efficiency. A sufficient condition for this to occur is exactly the strong bias condition $\varepsilon\theta > 1$.

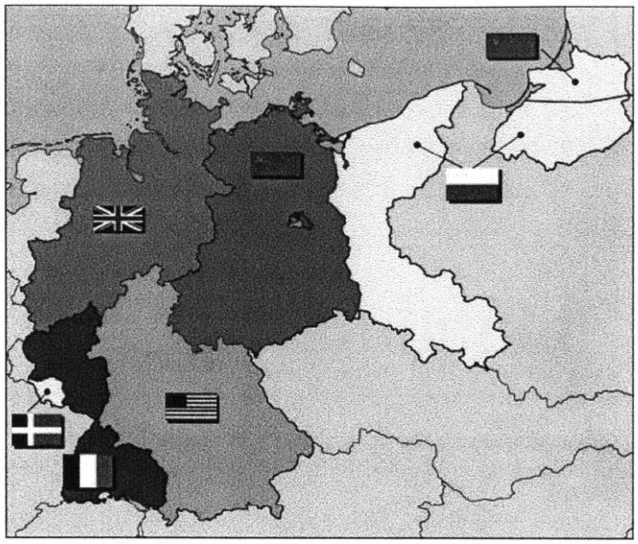
3 Empirical Analysis

Let me now turn to the empirical analysis. As already mentioned above, the analysis will have two broad parts, both of which are contained in Proposition 6 of the theoretical framework. Using individual-level microdata for multiple cross-sections, I will first study the differences in occupational choice and earnings between refugees and natives. According to the theory I predict both a higher employment share in unskilled occupations and lower earnings for refugees. Furthermore, the differences in earnings should stem to a large degree from differences in sorting, i.e. refugees earn less not because they are less productive, but because they work in relatively low-paying sectors and occupations. To test the additional predictions of the theory I will then turn to the experiences of different counties by exploiting the regional variation in refugee inflows. According to the theory, I predict that both wages in refugee-scarce sectors and employment shares in occupations requiring little specific experience are higher in refugee-rich counties. Furthermore, the data on wage-premia and average wages will be informative about the existence of biased technological adoption. Before I turn to the results however, let me briefly describe the historical background of the refugee inflows into Western Germany.

3.1 The Historical Background

After the second world war, Germany was not only divided into the four Allied Occupation Zones but also lost a large part of its Eastern Territories. The reason was that during the Potsdam Conference in 1945, Russia claimed to be entitled to the eastern territories of Poland as means of war reparations. As a compensation, Poland received the Eastern German territories on its western border. This regional reallocation is depicted in Figure 1. The map shows the four Allied Occupation Zones as the territory which comprises Germany today. The light shaded territories in the East are the regions which used to be German territories before the war, but which were allocated to Poland and Russia in the aftermaths of the war. In the Potsdam Agreement it was not only decided that those regions had to be integrated into Poland and Russia respectively, but also that the German population had to leave these regions and move to the German territories in the West. Augmenting this population transfer from regions that where previously German, the Potsdam Agreement also ruled that German minorities in Hungary and Czechoslovakia had to leave these countries to settle in Germany in its post war borders. The official protocol of the conference reads:

> "The Three Governments, having considered the question in all its aspects, recognize that the transfer to Germany of German populations, or elements thereof, remaining in Poland, Czechoslovakia and Hungary, will have to be undertaken.



This figure shows a map of Germany in 1945 after WW2. The light areas in the East, are the Eastern Territories, which were allocated to Poland and the USSR as means of war reparation. Western Germany is split into the different occupational zones under the control of France, UK, USA and USSR respectively.

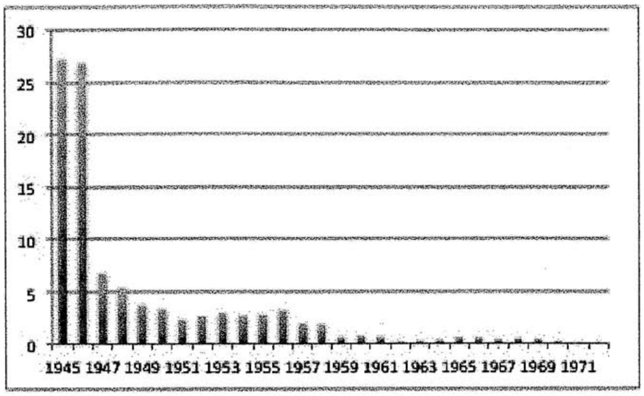
Figure 1: Germany after 1945

> They agree that any transfers that take place should be effected in an orderly and humane manner."

The subsequent population transfer is one of the largest transfers in world history. Between 1944 and 1950, roughly 12 million ethnic Germans were expelled and 8.5 million people were allocated to Western Germany (Reichling, 1958, p. 17). The timing of the arrival in Western German is depicted in Figure 2 below, which reports data from the German census.[3] It is clearly seen that the vast majority of the population transfer takes place during and in the two years immediately following the war. Most of people arriving in Western Germany after 1950 moved there following an initial spell in the Soviet Occupied Zone after their expulsion from the Eastern Territories.

The expulsion can be broadly divided into three phases. The first wave of refugees arrived in Western Germany during the last months of the war. Soviet forces made their appearance at the eastern German border in the summer of 1944. Trying to reach Berlin, soviet soldiers were advancing through the German Eastern Territories at great speed causing the German population to flee westwards. As the Nazi government considered the evacuation of German territories a defeatist act and executed a strict "no retreat" policy to use the civil population

[3] I will describe the data in more detail below.



Notes: The figure shows the distribution of expellees according to their time of arrival in Western Germany. For example 27% of the entire stock of expellees present in Western Germany in 1971 arrived in 1945. Source: *Mikrozensuszusatzerhebung 1971 (MZU 71)*.

Figure 2: Expellees' arrival in Western Germany

as a shield slowing down the Russian army, most inhabitants evacuated their homes fully unprepared. Because there were hardly any official evacuation plans as trains and ships were often reserved for the German soldiers, most refugees fled their homes by joining refugee treks, which suffered enormous casualties during the flight (de Zayas, 1993). After the German defeat in May 1945, the so-called wild expulsions started. These where mainly taking place in the spring and summer of 1945 before the Potsdam Agreement was signed in August 1945, most importantly in Poland and Czechoslovakia, where a substantial German minority resided. Under the backing of the respective governments, both the army and privately organized militias started to systematically expel the German population. It is only the Potsdam agreement which tried put an end to these unorganized expulsions and legalized them ex-post. Within the following two years, the majority of the German population was transferred from Middle- and Eastern Europe to Western Germany and the Soviet Occupied Zone. At the end of 1947, roughly 7 million people had arrived in Western Germany already. This amounted to roughly 20% of the population living in Western Germany at the time. Despite the casualties during the war, the population of Western Germany had therefore increased by 16% between 1939 and 1950.

3.1.1 The initial allocation of refugees in Western Germany

When thinking about the influx of refugees as a supply shock to labor markets, it is important how the initial allocation across the different regions in Germany was decided upon.[4] Both

[4]I will come back to the evidence on migratory responses below.

State	1946	1947	1948	1949	1950
Schleswig-Holstein	27	32	33	33	33
Hamburg	2	5	5	6	7
Lower Saxony	12	23	25	26	27
Bremen	4	5	6	7	8
Nordrhein-Westfalen	2	7	8	9	10
Hesse	3	14	15	16	16
Rheinland-Pfalz	1	2	2	3	3
Baden-Wuerttemberg	1	10	11	11	12
Bavaria	1	19	20	21	21

Source: Statistisches Bundesamt (1953, p. 5)

Table 1: Refugees as a share of the population in different states

the historical literature and the available government documents from the Official Military Government of the US in Germany (OMGUS) show very clearly that an orderly allocation of the refugees across localities in Western Germany was impossible at the time. The main reason was the administrative burden. While migration flows of this magnitude in such a sort time span would present a formidable challenge for any functioning state authority, the challenge in Germany in 1945 was enormous. Not only was the infrastructure in many places heavily destroyed through bombings, but the administration was also severely handicapped - at this time Germany did not even have a federal administration, but each occupied zone was governed by provisory military governments. One consequence of this situation was a striking degree of heterogeneity across states. Gerhard Reichling, a German historian who was in charge of the main statistical analysis about the integration of the refugees, for example concludes that "there is no aspect where the Federal Republic of Germany shows a similar degree of heterogeneity as in the absorption and distribution of expellees. The occupying forces decided about the contingents for their occupation zones without any German involvement....The fact that there is no central German state authority during the time of the expulsion and the missing solidarity of the states with respect to the expellees prevents an equitable and adequate distribution of burdens caused by the influx of expellees for years" (Reichling, 1958, p. 17). The wide disparity in the share of refugees in the state population can be seen in Table 1. Not only have three states (Bavaria, Schleswig-Holstein and Rheinland-Pfalz) by far the biggest refugee population share, but it is also seen how important geographic factors were for the initial allocation. As the official transports only started in 1946, it is only the expellees having flown during the first phase of wild expulsions and the war, which are living in Western Germany at this time. As the main escape route from the German Eastern Territories into Western

Germany was via the Baltic Sea, most refugees ended up in the northern states Schleswig-Holstein and Lower Saxony. Between 1946 and 1947, the official transports commence and Table 1 shows another sharp increase in the refugee shares. The main increase is found in Bavaria. Again geographical reasons seem prevalent as Bavaria has a common border with Czechoslovakia, where roughly 2 million German were expelled from between 1946 and 1947. But also Baden-Wuerttemberg, which has a common border with Bavaria and was also part of the US occupied zone, sees a substantial increase once the official transports began. Between 1947 and 1950, the distribution across states does not change drastically.

This lack of an orderly allocation mechanism had also important implications for local labor markets in that refugees were not allocated to regions, where they could make use of their existing skills. Werner Nellner, one of the leading post-war economic historians, describes the situation as follows: "In the midst of the chaotic post-war circumstances arrived the refugee transports. The entirely confusing political and economic situation paired with the abruptness of this pouring-in simply did not allow a sensible distribution of the expellees into areas where they could find work. The ultimate goal was to find shelter for those displaced persons, even though in the majority of cases the situation was very primitive and many had to dwell in the tightness of refugee camps for years" (Nellner, 1959, p. 73) and some observers even concluded that most refugees were "dumped into Western Germany and settled where they could" (Petersen, 1964, p. 420). That refugee flows were not aligned with labor market conditions was not only known to the military government, but was also considered an enormous problem at the time. As early as in 1946, P.M. Raup, Acting Chief of the Food and Agricultural Division of the OMGUS concludes that "both the planning and the execution of the support measures for German expellees was conducted entirely under welfare perspectives. The people in charge at the Military Government are social service officials. Similarly on the side of the German civil government, the department in charge is the social service agency. Entire communities are moved so that the population of some counties is increased by 25-30% and the agency in charge was founded to support the elderly, disabled people and the poor. Neither in Stuttgart, nor in Wiesbaden[5] we could find any information that either the Military Government or the German civil government even attempted to set up coordinated plans to receive, support or distribute the expellees. At no point, representatives of the Ministries of Labor, Industry, Nutrition or Education were consulted concerning the servicing or the distribution of expellees. The whole problem has not been handled as one of settlements of entire communities but as an emergency problem of supporting the poor." (Grosser and Schraut, 2001, p. 85). In a similar vain, a government official in Bavaria recalls that "there

[5]Stuttgart and Wiesbaden are the state capital of Baden-Wuerttemberg and Hesse, both states within the US occupied zone.

was no sensible planning whatsoever....The expelled peasants, worker and the industrialists, merchants, doctors, lawyers etc. had lost everything. Was there any chance that they could ever live independently again? That was their most pressing concern. The main problem for government officials and refugee commissars however was to provide provisional shelter during the winter of 1946/47" (Kornrumpf, 1979, p. 27).

The consequences on local labor markets of such lack of organization were already felt in the fall of 1947. In an official economic report by the OMGUS, it is argued that "expellees of Eastern Europe have been settled, not where their skills could be best utilized, but in accordance with the availability of food and housing to meet their needs. Thus, labor supply is often remote from the centers of labor demand ... Most of these have been placed without regard to their skills, and many where there is no demand or material for them to pursue their trades. The impact of these mass movements becomes evident when it is realized that one fifth of the population of the US zone has arrived there in the last seven and one-half years" (Office of the Military Government for Germany, 1947, p. 4-5).[6] With a special reference to industrial workers the report notes that "the industrial population, as pointed out previously, is not at all times resident in or near the labor market. Skilled laborers from the eastern territories are living in agricultural communities far from sources of employment. And even those who have been fortunate enough to find housing and food close to industrial centers have not always had the good fortune to find a demand for their special skills" (Office of the Military Government for Germany, 1947, p. 9). This excess supply of workers with the wrong skill set induced both job degrading and unemployment in the population of refugees. Julius Isaac of the National Institute of Economic and Social Research in London worked as a consultant for the OMGUS Civil Administration Division with a special reference to the problem of assimilation of refugees. He reports that "precious skilled labor has to work in unskilled occupations because there are no opportunities to work in their jobs.... Often they are located in rural areas where it is almost impossible to find appropriate employment" (Isaac, 1948). Furthermore, "practically the only unemployment among able-bodied male workers in western Germany, other than that of the white-collar group, is among the refugees. This is because the distribution of refugees was based on housing potentialities rather than on employment possibilities and the labor market." (Barton, 1948, p, 27).

3.1.2 Ex-ante heterogeneity between refugees and natives

As evidenced from this discussion, contemporaries at the time especially lamented the mismatch between skill-specific labor supply and demand for the refugees. So what are the most

[6]This citation refers to the population seven and one half years ago, as the last census in Germany was conducted in 1939.

important dimensions according to which refugees and natives differed in their formal skills and job-specific experience? A unique data appendix to the micro census in 1971 allows me to study the initial pre-expulsion heterogeneity, because it contains information on individual employment histories, i.e. individuals in 1971 were asked not only about their education level but also about their sectoral and occupational characteristics in 1939.[7] The results of this exercise are contained in Table 2, which compares expellees and natives according to various characteristics in 1939. I only consider individuals that were at least 18 years old in 1939, i.e. 50 or older in 1971. Panel A contains the educational characteristics of the two populations. It is seen that the distribution of formal skills was very similar. The only slight difference is the higher popularity of vocational schools (*Berufsschule*) in pre-war Western Germany, a fact that is due to the bigger importance of the manufacturing industry (see below). But overall, the population of refugees did not seem to be very different in terms of their formal skills. This is no longer true when we consider the sectoral structure of employment, given in Panel B. While the reliance on services and the public sector is very similar, there is a large disparity in the employment shares of manufacturing and agriculture. In particular, the agricultural sector is roughly 60% bigger in the Eastern Territories, with a commensurate smaller manufacturing sector. Finally, Panel C depicts the distribution across occupational classes, which again is very similar. Hence, the most important source of heterogeneity between the population of refugees and natives was the pre-war sectoral structure.[8]

3.1.3 Migration

As I am mostly interested in the long-term consequences of this historical episode, the migration response of initially misplaced workers is important. If we believe the assessment of contemporaries that the influx of refugees caused an initial excess supply in the labor market, where did the adjustment come from? While it is sensible that falling wages in refugee-rich labor markets should have triggered migration incentives, I will provide evidence that the initial supply shock in fact persisted for a long time. This was mainly due to severe restrictions on mobility in the late 40s and early 50s and small migration incentives afterwards. That neither individual migration nor organized population transfers within Western Germany had a large influence on the initial allocation in the late 40s was already seen in Table 1, which

[7] The only study that I am aware of that uses this data is Luettinger (1989).

[8] The sex-ratio is virtually identical in both population with females accounting for 59% of the population. Regarding the distribution of age, the population of refugees is slightly younger. This is mainly driven by the cohorts born between 1913 and 1919 being larger. As members of these cohorts are in their twenties in 1940, this difference could be driven by differential mortality during the war. However, given that the same pattern holds also true in the female population, I suspect that the difference is mostly due to lower birth rates during the First World War in Western Germany. As birth rates in war times not only drop sharply but especially so in urban areas, this is consistent with the fact that the Eastern Territories were more rural.

	Panel A: Educational Distribution			
	Refugees		Natives	
Elementary School	65.92		66.28	
High School	10.93		8.32	
Vocational School	15.53		18.42	
College	7.63		6.97	
N	36400		148194	
	Panel B: Sectoral Distribution			
	Males		Females	
	Refugees	Natives	Refugees	Natives
Agriculture	22.16	14.4	31.32	24.03
Manufacturing	43.12	52.75	24.32	29.88
Services	17.05	18.35	30.62	33.39
Public Sector	17.67	14.5	13.75	12.7
N	13,705	55,747	10,043	38,373
	Panel C: Sectoral Distribution			
	Males		Females	
	Refugees	Natives	Refugees	Natives
Self employed	9.57	9.47	5.13	5.53
Employees (skilled)	8.44	7.74	2.67	2.09
Employees (semi-skilled)	10.65	10.42	6.66	7.01
Employees (unskilled)	8.47	7.94	16.77	17.91
Workers (skilled)	2.88	3.61	0.7	0.51
Workers (semi-skilled)	25.89	26.63	5.09	5.14
Workers (unskilled)	21.84	23.84	36.56	36.69
Self employed (agricult.)	12.25	10.35	26.44	25.11
N	13,358	54,403	9,911	37,910

Source: The data stems from the *Mikrozensuszusatzerhebung 1971 (MZU 71)*.

Table 2: Heterogeneity in pre-war characteristics: refugees vs. natives

shows that the distribution of refugees is relatively stable across states until 1950. Despite the large heterogeneity in the economic burden the refugees imposed on individual states, there were hardly any organized transfers. The main reason for this is was the reluctance of states to accept refugees and it was only after the foundation of the Federal Republic of Germany in 1949, that there was enough political pressure for a centralized solution for the redistribution to be found. Before 1950 however, "all proposals for specific reallocation rules were rejected. The main argument was that the available data about the housing situation was not comparable across states. Hence, it was demanded that a general housing census was needed, which was only conducted in 1950 in conjunction with the general census. ... Measures concerning a population redistribution within the area of Germany had only a minor influence until the end of 1950" (Nellner, 1959, p. 39, 43).[9] It is this census which I use as my source of information about the county level allocation of refugees when I study the long-run impacts on local labor markets below.

To understand why expellees did not respond to the aforementioned adverse conditions on the labor market with out-migration, it is important to note that labor mobility was severely restricted in the post-war period. In 1945, the refugee committee of the Occupying Forces decided to deploy armed forces at the state boundaries to prevent internal migration (Fluechtlingsausschuss des Laenderrats (1945)) and William H. Draper, Director of the Economic Division of the OMGUS, notes that "Germany has been virtually cut into four Zones of Occupation - with the Zone borders not merely military lines, but almost air-tight economic boundaries" (Office of the Military Government for Germany, 1945, p. 10). Additionally, the incentives to migrate were also arguably low. In the late 40s, Germany was still characterized by a severe housing shortage both due to the destruction of the war and the population increase through the expellees. The majority of expellees either lived in refugee camps or with the native population who were forced by the Military Governments to accept expelled families in their homes. This political support however was only provided for individuals being part of the official refugee transports. Similarly, there were restriction to receive food stamps without being officially registered (Grosser and Schraut, 2001, p. 83).[10] Finally, the economic reports of the OMGUS themselves argue that high levels of unemployment are accompanied by labor shortage because "the mobility of labor is limited. Hence there is little possibility of an early change in the distribution of labor. For example. 46% of the job openings in Bavaria

[9] It is hard to tell if this was a real problem or if those states that were about the receive expellees from Bavaria, Schleswig-Holstein and Lower Saxony used this argument as an excuse to postpone large scale population transfers.

[10] Additionally the Allied governments were very reluctant to allow for free migration as they were afraid that expellees would form non-integrated sub populations along former village or city lines. To achieve this, various organizations trying to track down friends and family members were forbidden until the early 1950s (Nellner, 1959, p. 75).

State	1950	1962	1963	1964	1965	1966	1967	1968	1969
Schleswig-Holstein	33	28	29	27	26	27	27	27	26
Hamburg	7	12	11	10	11	12	11	11	10
Niedersachsen	27	25	25	24	24	24	24	23	24
Bremen	8	15	14	15	15	14	14	14	14
Nordrhein-Westfalen	10	15	13	14	13	13	13	14	13
Hessen	16	17	17	17	17	17	16	16	16
Rheinland-Pfalz	3	8	8	8	8	8	8	7	7
Baden-Wuerttemberg	12	16	16	16	16	15	14	14	14
Bavaria	21	19	20	18	18	18	17	17	16
Correlation with 1950		0.98	0.99	0.97	0.97	0.98	0.98	0.98	0.97

Source: The table shows the distribution of refugees as a share of the respective state population in different years. The data for the 1960s stems from the annual census from 1962 - 1969. For the data for 1950 see Table 1.

Table 3: Distribution of refugees across states in the 1960s and 1950

in March 1947 were in the major cities of Munich, Nuremberg and Augsburg, while the majority of immigrant labor resided in rural districts. In consequence, the economic absorption of immigrants is greatly hampered" (Office of the Military Government for Germany, 1947, p. 10).

While such administrative constraints seemed to have been loosened in the 1950s, the census data in the 1960s contains information on the state of residence of refugees. This data, together with the distribution in 1950, is depicted in Table 3. While the wide disparity between states 1950 is somewhat lower in 1960s, the initial refugee allocation is still very much visible. As for the initial allocation, Schleswig-Holstein, Lower Saxony and Bavaria still have the highest refugee share 15 years later and the distribution looks very stable in the 60s. Compared to the 50s, mainly Nordrhein Westfalia and Rheinland Pfalz (both of which share a border with France) have increased their population share of expellees. This is probably partly due to internal migration and partly due to the relatively small but continuing inflow of refugees through the 50s and 60s (see Figure 2). Overall however, Table 3 shows that the initial supply shock of refugees persisted (at least at the state level) for the later decades - as the final row in the column shows, the correlation for each year with 1950 is almost unity.

As a final piece of evidence, consider Table 4, which shows the share of refugees in cities of different size categories.[11] Refugees are significantly underrepresented in large cities with more than 100.000 inhabitants and more likely to live in relatively rural areas.[12] Again, this

[11]This data is available for each year between 1962 and 1969. To save some space I only report these two years. As can be expected, the results are very similar for the other years.

[12]That the share of refugees in very small settlements with less than 2000 inhabitants is relative low, should not come as a surprise. Arguably, these villages are very reliant on agriculture, a sector where expellees with

		Less < 2000	2,000 to 10,000	10,000 to 100,000	more than 100,000
1962	Refugees	22.85	26.67	27.86	22.63
	Natives	25.3	21.83	23.93	28.95
	N	32613	29672	32221	36544
1969	Refugees	15.43	24.93	31.93	27.72
	Natives	19.73	20.53	25.04	34.69
	N	88971	98887	121639	156847

Source: The data stems from the annual census in 1962 and 1969.

Table 4: Share of natives and refugees in cities of different sizes

is consistent with limited migration within state-boundaries as the initial allocation after the war was biased towards rural areas as the shortage of housing was even more severe in cities due to war destruction.

3.2 Post-Expulsion Outcomes

Let me now turn to the long-term outcomes of this historical episode. Before doing so, I give a brief description of the various data sources used.

3.2.1 Data

This paper uses various data sources from the 60s and 70s. The most important dataset is the *Mikrozensus Zusatzerhebung 1971 (MZU 71)*, a special appendix to the census conducted in 1971. The purpose of this dataset was to study the "social and economic mobility of the German population" and fortunately it includes identifiers about individuals' refugee status. Most importantly, the data contains retrospective information about employment characteristics in 1939, 1950, 1960 and 1971 at the individual level. This allows me to observe the whole employment history of individuals and hence distinguish cohort from life-cycle aspects. The MZU 71 has roughly 200.000 observations, 40.000 of which are refugees. The MZU 71 data does not contain information about wages. Hence, to study the implications on factor prices, I use two additional micro datasets that contain information on both wages and the refugee status of respondents. The *Einkommens-und Verbrauchsstichprobe 1962/63 (EVS 62)* is a micro dataset conducted in 1962 to measure household income and expenditure and is hence similar to the Consumer Expenditure Survey in the US. The 1962/63 wave of the survey has about 32.000 observations. Additionally I use the *Volkszaehlung 1971 (VZ 71)*, i.e. the census. Having two wage measures two decades apart turns out to be interesting, because

their lack of land titles simply could not enter.

we can compare how cohort effects affect the relative wages between natives and refugees. The VZ 71 is a dataset with roughly 150.000 observations. Finally, the time series about the distribution of refugees across states shown in Table 3 is taken from the annual census data from 1962-1969, each of which has between 420.000 and 480.000 observations.

To measure the effect of refugee inflows on local labor markets, I use the 1975 cross-section of the regional files of the IAB employment samples (IABS).[13] The IABS consists of a 2% random sample of social security records and is for example also used in Dustmann, Ludsteck, and Schoenberg (2009). The important property of the regional files of the IABS is that it contains regional identifiers about the *Arbeitsmarktregion*, the local labor market. I will be referring to those regions as a county. There are roughly 250 counties in Germany and the data is aggregated so that each county has at least 100.000 individuals. This data is matched to the initial allocation of refugees in 1950, which I coded from Statistisches Bundesamt (1955). According to Figure 2, by that time roughly 80% of refugees have arrived in Western Germany. Unfortunately, some county boundaries were subject to redistricting in the 50s and 60s so that the mapping between counties in 1950 and 1975 is noisy. I built a crosswalk using the official information, but the redistricting clearly introduced some measurement error in the county-specific refugee measures.

Finally, I use the 1985 qualification surveys conducted by the IAB and the BBIB that provide direct evidence on the type of technologies used on the factory floor.[14] These surveys, which are also used in Spitz-Oener (2006) and DiNardo and Pischke (1997), aim to measure the required skills and technologies employed in different detailed job categories. They contain detailed information on which technologies workers use on a day-to-day basis, if their educational background is important for their job or if most skills are acquired through learning-by-doing. The survey samples 30.000 workers and I can match the BBIB Qualification Data and the IABS regional files on the level of approximately 80 job categories.

3.2.2 Individual-level results

According to the historical sources, the initial allocation of refugees caused a mismatch between skill-specific labor supply and demand on local labor markets. The way equilibrium in the labor market was restored was through a massive sectoral reallocation of refugee labor into the manufacturing industry. To gauge the quantitative magnitude of this reallocation, consider Table 5, where I show the cross-sectional results of the employment histories contained in the MZU. The table contains the pattern of sectoral employment for 3 decades for the two

[13]The IAB is the Institute for Labor Market Research at the German Department of Labor.

[14]The BBIB is the Federal Institute for Vocational Education and Training, a federal government institution for policy and research in the field of vocational education and training.

	1939		1950		1960		1971	
	Refugees	Natives	Refugees	Natives	Refugees	Natives	Refugees	Natives
Agriculature	26.04	18.7	12.63	17.43	4.92	12.81	2.49	9.99
Manufacturing	35.87	43.64	51.92	46.53	56.65	47.85	52.01	45.26
Services	23.2	24.77	20.96	23.94	22.01	25.23	24.75	26.01
Public	14.89	12.89	14.49	12.11	16.43	14.12	20.76	18.74
N	27,618	106,727	34,945	141,811	46,005	191,264	48,299	200,779

Source: The data stems from the *Mikrozensuszusatzerhebung 1971 (MZU 71)*.

Table 5: Employment shares in different sectors: natives vs. refugees

respective populations. For comparison, the first two columns contain again the sectoral employment patters in 1939 (and is therefore simply the full population equivalent of Panel B of Table 2), which showed a significantly larger manufacturing sector in Western Germany. Columns thee and four contain the results of the post-expulsion labor market equilibrium, i.e. after roughly 8 million refugees arrived in Western Germany. Focusing on the refugees first, we see a drastic example of sectoral mobility. Essentially 50% of workers who used to be employed in the agricultural sector are now working in the manufacturing industry.[15] However, when we look at the experiences of natives, we do not see *any* sectoral mobility in that the sectoral employment patterns between 1939 and 1950 are almost unchanged. As Table 2 showed that refugees and natives did not differ much in their formal skills (as measured by education), this heterogeneity in outcomes has to stem from pure experience effects.[16] Through the lens of the theoretical framework, I interpret these differences in the occupational choices of natives and refugees as differences in comparative advantage.

In fact, the official documents at the time discuss exactly this reallocation process. In an early report from the US Military Government it is argued that "a clue of future developments, however, lies in the study of the character of the migration. Of the immigrants who came to the US Zone from east of the Oder-Neisse, 19.6%, or well over half a million, were farmers. But agricultural acreage, already reduced as a consequence of four-power reductions of the national territory, cannot be expanded significantly. Within the US Zone the possibility of increasing settlement by changing the size and structure of farms is very small" (Office of the Military Government for Germany, 1947, p. 8).[17] "As a result both of the necessity to

[15] Table 2 contains the results for the entire population. However, the results are not driven by composition effects in that for example female labor market participation dropped. The same results holds for both males and females.

[16] Another explanation would obviously be outright discrimination. In the historical literature discrimination is indeed mentioned, but it is hard to quantify how widespread such discriminatory attitudes were. Recall also that this is an example of migration flows, where migrants share the same culture and language.

[17] The difference between the 19.6 percent farmers mentioned in this government report and the slightly higher number reported in Table 2 probably stems from the fact that in the US Zone there were also immigrants from

work and of the unfavorable supply-demand relations in the refugee labor market, relatively high numbers of refugee wage and salary earners are forced to work in occupations other than those they have learned or in which they would normally be employed" (Office of the Military Government for Germany, 1947, p. 34).

The high relative elasticity of labor demand for inexperienced occupations, which in the theory is captured by the degree of increasing returns (θ), can also be seen in the data on the occupational distribution, which is depicted in Table 6. Again, the first two columns contain the data for 1939 and is therefore the full population equivalent to Table 2. Turning to columns three and four we see that the occupational distribution is relatively stable except for the case of unskilled workers, most of which work in the manufacturing industry. In the population of refugees the employment share in unskilled occupations surges by 50% relative to its pre-expulsion level. Again, there is hardly any change in the occupational distribution for natives - if anything, the share of both unskilled workers and employees falls slightly. If skilled and unskilled labor inputs are complements this is consistent with refugees working in unskilled occupations and natives experiencing occupational upgrading as a response. Together, Tables 5 and 6 paint a clear picture of the short run adjustment of the local labor market: Most of the burden of adjustment - both at the sectoral and the occupational level - fell on the refugees and with roughly 25% of the population being directly affected, the magnitudes are very large.

While it is tempting to interpret these findings as job-mobility within workers' life-cycle, Tables 5 and 6 are cross-sectional facts. Hence, they might reflect mainly composition effects, where most sectoral (and occupational) mobility takes place across different cohorts. The unique feature of the MZU data allows us to rule out this possibility, because the data stems from individual employment histories, i.e. the information is essentially a 30 year panel. In Figure 3 I depict the employment share in manufacturing for different cohorts for both refugees and natives. Three aspects stand out. First of all, for the refugee population, the manufacturing share is strongly increasing between 1939 and 1950 for all cohorts. Hence, the sectoral mobility in the cross-section reflects sectoral change within individual employment histories. Secondly, in both populations we see strong cohort effects, in that younger cohorts have much higher employment shares in manufacturing. This of course reflects the secular transition from agriculture to manufacturing.[18] Finally, the refugee-native difference is very visible. Not only do refugees "leapfrog" their native counterparts after the expulsion, but while there is still a strong increase in manufacturing employment for refugees, the profile for natives

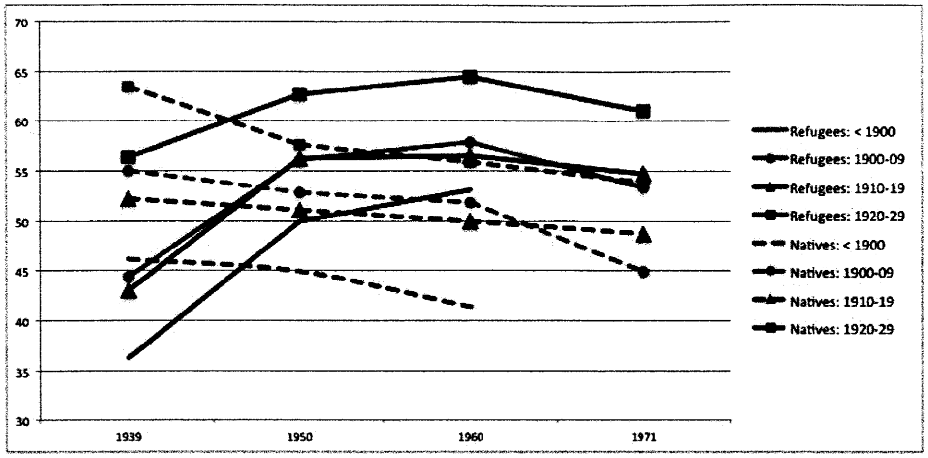
Czechoslovakia, which was less reliant on agriculture.

[18]In the Appendix I also show the respective cohort profiles for the employment share in agriculture. Juxtaposing the results for manufacturing, it shows marked cohort effects (with younger workers being less reliant on agricultural employment), a secular decline over time and refugees leaving the agricultural sector after the expulsion

	1939		1950		1960		1971	
	Refugees	Natives	Refugees	Natives	Refugees	Natives	Refugees	Natives
Self employed	6.98	7.29	4.9	8.82	4.99	8.5	5.51	8.71
Employees (skilled)	5.44	5.04	5.33	5.42	6.99	6.97	9.63	9.6
Employees (semi-skilled)	8.4	8.57	8.15	9.23	11.34	10.95	16.06	14.97
Employees (unskilled)	12.3	12.15	10.29	10.53	11.85	11.78	12.96	12.4
Workers (skilled)	1.88	2.21	2.01	2.4	1.9	1.92	2.55	2.28
Workers (semi-skilled)	16.44	17.38	19.35	19.12	22.55	20.24	19.06	17.43
Workers (unskilled)	29.76	30.31	46.34	27.86	37.54	26.07	31.57	23.29
Self employed (agricult.)	18.8	17.05	3.62	16.61	2.84	13.58	2.65	11.33
N	25,745	100,065	32,541	132,630	43,108	176,599	44,893	188,493

Source: The data stems from the *Mikrozensuszusatzerhebung 1971 (MZU 71).*

Table 6: Employment shares in different occupations: natives vs. refugees



Notes: The figure shows the employment shares in manufacturing for different cohorts in both the refugee and the native population. The results for refugees are shown in solid lines, the results for natives are the dashed lines.

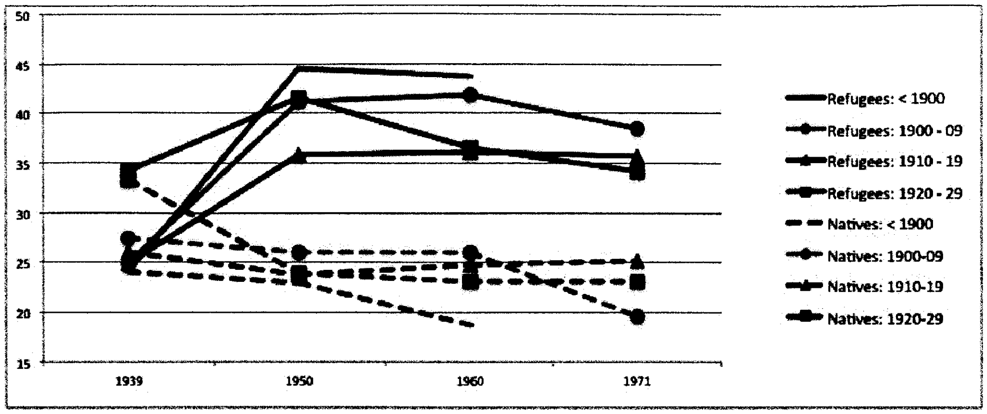
Figure 3: Employment share in manufacturing for different cohorts

is essentially flat.[19] Hence, the life-cycle evidence within cohorts paints the same picture as the cross-sectional date: the magnitude of sectoral reallocation is large and takes place within individual employment histories. That the same holds true for occupational mobility is seen in Figure 4. While the employment shares in unskilled occupations are almost identical *for each cohort* in 1939,[20] there is a sharp increase in unskilled employment for the refugee population after the expulsion.

While the immediate effect of the expulsion in the mid 40s is probably the most interesting pattern, it is also instructive to look at the long-run developments during the 60s and 70s. As evidenced from Table 5 and Figure 3, the initial "push" of the refugee population into the manufacturing sector was very persistent, i.e. even in 1971, the employment share in manufacturing exceeds the one of the native population significantly. The time-series behavior of the occupational distribution is slightly different however. As seen in Figure 4, like for the case of sectoral employment patterns, there is very little occupational mobility within the life-cycle for older cohorts, i.e. individuals born before 1920. For them, the initial job-assignment in the late 40s seems to persist through their entire working careers. This is different for younger cohorts. The generation born after 1920 seems to experience some occupational mobility during its life-cycle as the share of employees in unskilled occupations shows a downward

[19]The slight negative trend stems probably from sector-specific age profiles, where both entry in and retirement from the manufacturing sector occurs earlier in the life-cycle.

[20]Recall that Table 2 already showed that occupational patterns were not much different in the pre-war era. However, Table 2 did not stratify the sample by cohorts.



Notes: The figure shows the employment shares in unskilled occupations for different cohorts in both the refugee and the native population. The results for refugees are shown in solid lines, the results for natives are the dashed lines.

Figure 4: Employment share in unskilled occupations for different cohorts

trend. However, relative to the generation of natives, unskilled occupations remain much more prevalent in the population of refugees. Figure 4 therefore implies that decrease in the share of unskilled workers in the cross-section of refugees is mostly driven by cohort effects, as e.g. models of different skill vintages imply (Chari and Hopenhayn, 1991).

Let me now turn to the wage implications of this reallocation. According to the theory, refugees earn lower wages than their native peers because of their lack of opportunities in some sectors of the economy. Hence, earnings in the population of refugees are lower, not because of their sector-specific skillset, but because of them having to sort into industries, which in equilibrium end up paying lower wages. Unfortunately there is no microdata available for the immediate post-war period. So the earliest existing data source is the EVS in 1961. While the original purpose of this dataset was to study individuals' expenditure patterns, it also contains information on earning and job characteristic. In Table 7 below I report the results of a simple decomposition of earnings of the form

$$ln\left(w_i\right) = \alpha + \beta Ref_i + \gamma Ref_i \times Young_i + \delta_s + \delta_o + X_i'\eta + u_i, \tag{15}$$

where w_i denotes individuals $i's$ earnings, Ref_i is an indicator variable of the refugee status, $Young_i$ is an indicator if the individual's age is below 35 (so that the individual was 20 at the time of the expulsion and hence started his or her working career in Western Germany), δ_s and δ_o are sector and occupation fixed effects and X_i contains additional controls. According to column one in Table 7, 15 years after the initial expulsion refugees still earn about 10% lower wages. Column 2 shows that this effect is only half as big for refugees that started their labor

	Dependent variable: $ln(w)$				
Refugee	-0.0924**	-0.102**	-0.103**	-0.0719**	-0.0356**
	(0.00820)	(0.00957)	(0.00953)	(0.00897)	(0.00841)
Refugee × Young		0.0453**	0.0432**	0.0188	0.0163
		(0.0161)	(0.0160)	(0.0147)	(0.0141)
Citysize FE	No	No	Yes	Yes	Yes
Sector FE	No	No	No	Yes	Yes
Occupation FE	No	No	No	No	Yes
N	27506	27506	27506	27506	27506

Notes: Robust standard errors are shown in parentheses. ** and * denotes significance at the 5% and 10% level respectively. All regressions include a set of state fixed effects, a quadratic age profile and a dummy for respondents' sex.

Table 7: The wages for refugees in 1961

market career in Western Germany and hence are arguably less prone of being disadvantaged in some sectors of the economy. Column 3 introduces a set of citysize fixed effects as Table 4 above showed that refugees were more likely to reside in rural areas. This has no effect on the results. Finally, columns 4 and 5 are informative about the origins of these lower wages and hence about the mechanism stressed by the theory. By introducing sector fixed effects, the refugee coefficient drops by 30% and there is no longer a significant differential for younger workers. Hence, 30% of the wage difference between refugees and natives is due to the former working in low-wage sectors and this being especially so for refugees that were expelled while being already in the labor force. Finally, column 5 introduces an additional set of occupation fixed effects, which reduces to coefficient to 35% of its initial value. Hence, 15 years after having to integrate in the western labor market, refugees still earn substantially less than their Western counterparts and almost 70% of this difference is due to the fact that they ended up in low-wage sectors and occupations.

The other micro dataset, which allows me to study the earnings of refugees is the census in 1971 (VZ 71). Compared to the EVS it has the advantage of also containing information on individuals education levels. The results of running a specification akin to (15) are contained in Table 8 and line up very well with the ones from a decade earlier. Column 1 shows that refugees only earn 5% less then their peers, which I interpret as the result of the most severely disadvantaged cohort having left the labor force. Columns 2 to 4 show that controlling for educational attainment and sector leaves the refugee coefficient essentially unchanged - especially the insensitivity when controlling for education is reassuring, because there are little educational differences between the two populations. Finally, column 5 controls for a

	Dependent variable: $ln(w)$				
Refugee	-0.0586**	-0.0577**	-0.0587**	-0.0567**	-0.0250**
	(0.00295)	(0.00271)	(0.00272)	(0.00269)	(0.00232)
Education FE	No	Yes	Yes	Yes	Yes
State and citysize FE	No	No	Yes	Yes	Yes
Sector FE	No	No	Yes	Yes	Yes
Occupation FE	No	No	Yes	No	Yes
N	150392	150392	150392	150392	150392

Notes: Robust standard errors are shown in parentheses. ** and * denotes significance at the 5% and 10% level respectively. All regressions include a quadratic age profile and a dummy for respondents' sex. Education is measured by 7 educational categories (elementary school, three types of vocational school, high school degree, degree from an engineering school and college degree).

Table 8: The wages for refugees in 1970

set of occupational effect, which expectedly reduces the coefficient by another 50%. Hence, the general pattern is similar to the results in 1961, albeit at a lower level. These results are consistent with the theoretical mechanism, which stresses occupational choice as the main determinant of earnings. Not only do occupational fixed effects account for the largest part of the wage differences between natives and refugees, but the average wage penalty also declines over time as older cohorts with their pre-expulsion skill set reach the stage of retirement.

3.2.3 County-level results

After studying the individual data on wages and sectoral mobility, I now turn to the local labor markets itself and exploit the variation in the allocation of refugees across counties. Consider first individual wages. I am interested to estimate a reduced form relationship of the form

$$ln\,(w_{ic}) = \alpha + \beta s_c^{Ref} + X_{ic}'\eta + u_{ic}, \tag{16}$$

where w_{ic} denotes the hourly wage of individual i in county c, s_c^{Ref} is the share of refugees allocated to county c in 1950 and X_{ic} denote individual and county-specific controls. According to the theory, β corresponds to the effect of refugees inflows on average wages or county welfare. There are two main challenges to estimate (16). The first one simply concerns the availability of data. The earliest microdata with regional identifiers to match individual workers to a level of aggregation lower than the state level (of which there are only 9 in Western Germany) dates from 1975 and is part of the IABS. The second problem is the endogeneity of the initial allocation of refugees across counties. The historical evidence reported above already showed that refugees were mostly sent to rural areas outside larger cities. So while the composition of refugees was arguably exogenous with respect to county characteristics, the level of refugees was surely not. To account for this endogeneity I will use prewar information as an instrument. According to the historical sources, the main determinant of refugee flows were housing and food supplies. Hence, I will estimate β in (16) by instrumenting s_c^{Ref} with the county-level population density in 1939 and including the population density in 1975 as a control in X_{ic}. The identification assumption is that conditional on the population density in 1975, the population density 35 years earlier has no effect on individual wages (other than through being a determinant of the allocation of refugees). There are two main worries why this exclusion restriction might be violated. The first one concerns war destruction. As high population density counties were more affected by the bombing of the allied forces, war related damages are likely to be correlated with the initial population density in 1939. If such damages affect wages conditional on the current population density, the exclusion restriction will fail. While there is evidence regarding the dynamics of citysize that war related damages have little long-lasting effects (Davis and Weinstein, 2002; Brakman, Garretson, and Schramm, 2004), it is clearly a possibility that wages are affected in the long-run. The second worry concerns the importance of dynamics. With the population density in 1975 being a control in the regression, the exclusion restriction also implies that (conditional on the current level of population) population growth in the past does not affect wages. Again, there are reasons to believe why this restriction might fail, but given the data at hand, past population density is best instrument at my disposal.[21] In Table 9 I report the results of estimating a version of (16) in the cross-section of workers in 1975. In particular, I consider a specification of

$$ln\left(w_{ic}\right) = \alpha + \beta s_c^{Ref} + \gamma s_c^{Ref} \times J_i + \tau ln\left(popdens_c^{75}\right) + X_{ic}'\gamma + u_{ic}, \tag{17}$$

where s_c^{Ref} is the (instrumented) refugee share, J_i are different individual-specific characteristics to study the differential effect of the initial refugee allocation on particular subgroups of the population and X_{ic} contains an education-experience profile and a full set of sector fixed effects, which I allow to vary for different educational groups. According to the theory, I expect β to be positive if there is strong bias, i.e. if $\varepsilon\theta > 1$. Column 1 first presents the simple OLS without any county-specific controls. The negative coefficient on the refugee share reflects the adverse selection of counties receiving large refugee inflows, i.e. high-refugee counties in 1939 were more rural and still have lower wages in 1975. In the theoretical framework, this reflects a negative correlation with the local productivity measures A or φ. Columns 2 adds the current population density as a control, i.e. column 2 is the OLS version of (18). Accounting for

[21] Another alternative, which I will pursue in the future is to use distance measures to the Eastern Territories as an instrument for refugee flows. Casual inspection of the maps suggests that counties close to the eastern border received much more refugees than counties on the French border.

	Dependent variable: $ln(w)$					
	OLS		IV			
share of refugees (s^{Ref})	-0.260*** (0.0732)	0.141*** (0.0520)	0.538*** (0.161)	0.603*** (0.182)	0.501*** (0.182)	0.732*** (0.179)
log population density (1975)		0.0336*** (0.00239)	0.0463*** (0.00537)	0.0482*** (0.00708)	0.0478*** (0.00706)	0.0480*** (0.00710)
empl share agricult. (1939)				-0.0306 (0.0443)	-0.0301 (0.0441)	-0.0326 (0.0440)
empl share industry (1939)				-0.0770** (0.0364)	-0.0747** (0.0364)	-0.0777** (0.0365)
$s^{Ref} \times Young$					0.198*** (0.0340)	
$s^{Ref} \times MedEduc$						-0.170*** (0.0464)
$s^{Ref} \times HiEduc$						-0.151 (0.114)
N	201607	201396	194478	193550	193550	193550
R^2	0.208	0.216	0.213	0.212	0.216	0.211

Notes: Standard errors are clustered on the county level and shown in parentheses. There 259 clusters. ***, ** and * denotes significance at the 1%, 5% and 10% level respectively. All regressions include a quadratic experience profile, a set of state fixed effects and a full set of education (3 categories) and sector (16 categories) interactions. For the instrumental variable estimates in columns 3 to 7, the excluded instrument is "log population density in 1939" at the county level. In the first stage regression, "log population density in 1939" has a t-statistic of -6.95.

Table 9: The effect of refugees on county level wages in 1975

the strong negative correlation between the initial refugee share and the population density in 1975 turns the coefficient positive.[22] The point estimate implies that a 10%-point increase in the share of refugees increases average wages by roughly 1.5%. Columns 3 contains the instrumental variable estimates. This increases both the coefficient and the standard error substantially. This increase is probably to a large degree due to attenuation bias in the OLS estimates as the construction of the county-specific refugee shares is prone to mismeasurement given the redistricting of the county boundaries (see the discussion above). Part of the increase however can also stem from the correlation between the refugee allocation and the necessity of food supplies. Holding the population density fixed, counties with a larger of agricultural sector were allocated more expellees. If those counties have local labor markets where average wages tend to be low, we expect the OLS coefficient to be downward biased. Column 4 introduces two coarse measures of the pre-war industrial structure, namely the agricultural and the industrial employment share (with the share of people working in services and the public sector excluded). While this does not change the refugee coefficient, it is surprising that the share of industrial workers in 1939 is negatively correlated with average wages while the agricultural share is not (at least not significantly so).[23]

In the remaining columns I now let refugees have differential effects for different subgroups in the labor market. This is a first pass to test for the existence of biased technological adoption. Table 9 shows that wages in refugee-rich counties are especially high for young workers (columns 5) and for relatively unskilled workers. If these sub populations have relatively high employment shares in occupation, which are not experience intensive (in terms of the model, the factory sector), these results are consistent with endogenous bias. At this level of aggregation however, it is of course impossible to identify the exact mechanism of these differential effects. To make progress on this dimension, I therefore now turn to the detailed information on job-characteristics contained in the BBIB data. The theory predicts that firms in refugee-rich counties should have adopted technologies, which require little prior experience and do not rely on human capital that is technology-specific. The BBIB data allows me to at least construct crude measures for such job-characteristics. First of all, it contains data on the origins of job-specific knowledge, i.e. are the skills required for the particular job mainly acquired via leaning on the job or formal education. Secondly, it asks individuals what type of technologies they regularly use and these technologies are be classified as skill- and unskill biased. As an example consider Table 10, which contains some technologies classified as either skill and unskill-biased by the survey. Arguably, refugees had (on average) a comparative

[22] In a regression of the refugee share on log population density in 1975 and a set of state fixed effects, the coefficient is -0.025 with a t-statistic of 8.

[23] Note however that the 1939 employment share in the service sector is the excluded category, which is higher in urban areas.

Unskill biased	Skill biased
Simple tool	Computer
Manual measuring tool	Numerical controlled machine
Hand drill	Chemical industry machine

Table 10: Skill- and unskill biased technologies in the qualification surveys

advantage in the usage of hand-drills relative to the use of numerically controlled machines. In the appendix I also report additional checks for the internal validity of these measures and show that educational attainment, individual earnings and firmsize are negatively correlated with the usage of unskill-biased technologies and with the importance of learning-by-doing. If the inflow of refugees triggered the adoption of technologies that were unskill biased and could be learnt on the job, wages in refugee-rich counties should be relatively higher in those jobs. In terms of the theory, this is precisely a test of the strong form of bias, i.e. if $\varepsilon\theta > 1$. To test this hypothesis, I consider a specification of the form

$$ln\left(w_{ijc}\right) = \alpha + \beta s_c^{Ref} + \gamma s_c^{Ref} \times Bias_j + \tau ln\left(popdens_c^{75}\right) + X_{ic}'\gamma + u_{ijc}, \tag{18}$$

where $Bias_j$ is the respective measure of the bias of technology. In addition to the learning-on-the-job and the technology measure described above, I also construct two additional measures of the skill content, namely the share of unskilled occupations and share of unskilled individuals in each job. The theoretical counterpart of the parameter γ in (18) is exactly the wage premium in the factory sector. The results are contained in Table 11 and are consistent with the existence of a strong form of technological bias. In particular, refugee-rich counties have especially high wages in jobs where learning on the job is important, where unskilled occupations are predominant, where the share of unskilled individuals is high and where unskill-biased technologies are employed. When interpreted through the theory, this is indicative of a large degree of increasing returns and is hence consistent with the positive effect on average wages reported in Table 9 and the high demand elasticity in unskilled occupations (Table 6).

Proposition 6 showed that such biased demand shifts should not only be reflected in factor prices, but also in region-specific employment shares. In particular, refugee-rich counties should have relatively high employment shares in unskill-intensive factory jobs. To study this implication, I simply calculate county $c's$ employment share in sector or job j, ρ_c^j, from the microdata and run a specification of the form

$$\rho_c^j = \alpha + \beta s_s^{Ref} + \tau ln\left(popdens_c^{75}\right) + X_c'\gamma + u_c, \tag{19}$$

	Dependent variable: $ln(w)$			
share of	0.171	0.393**	0.452**	0.574***
refugees (s^{Ref})	(0.200)	(0.198)	(0.204)	(0.191)
$s^{Ref} \times$	0.876***			
Learning On Job	(0.0911)			
$s^{Ref} \times$		0.916***		
Unskilled occup.		(0.0947)		
$s^{Ref} \times$			0.246**	
Low educational attain.			(0.109)	
$s^{Ref} \times$				0.132
USB technology				(0.0989)
Observations	189485	189485	189485	189485
R^2	0.232	0.250	0.296	0.246

Notes: Standard errors are clustered on the county level and shown in parentheses. There 259 clusters. ***, ** and * denotes significance at the 1%, 5% and 10% level respectively. All regressions include a quadratic experience profile, a set of state fixed effects, a full set of education (3 categories) and sector (16 categories) interactions and control for the log of the population density in 1975 and the agricultural and the industrial employment share in 1939. The excluded instrument is "log population density in 1939" at the county level and the respective interactions.

Table 11: The biased effects of refugee inflow

where I again instrument the refugee share with the population density in 1939. The results are contained in Table 12. The first two columns contain the results for the size of the manufacturing sector. Reassuringly, the initial allocation of refugees is highly positively correlated with the county-specific employment share in manufacturing. Given the evidence on the individual level, this should not be too surprising. Note however that the coefficient implies a larger increase in the manufacturing share than simply an "absorption" of the refugees. Letting $\rho_{c,R}^{M}$ and $\rho_{c,N}^{M}$ be the employment share in manufacturing for refugees and natives respectively, we of course have

$$\rho_c^M = \rho_{c,R}^M s_c^R + \rho_{c,N}^M \left(1 - s_c^R\right) = \rho_{c,N}^M + s_c^R \left(\rho_{c,R}^M - \rho_{c,N}^M\right). \tag{20}$$

While refugees are more concentrated in the manufacturing sector (i.e. $\rho_{c,R}^{M} > \rho_{c,N}^{M}$), this difference is clearly smaller than 1, which is the coefficient reported in Table 12. This suggests, that the inflow of refugees triggered job-creation in the manufacturing industry, which caused also the manufacturing share of natives to be higher in refugee-rich counties - in the notation of (20), $\rho_{c,N}^{M}$ is positively correlated with s_c^R. While this is clearly possible, the reduced form effect still seems very high. The estimate of the elasticity, contained in column 2, is more reasonable. Empirically, the both the mean refugee share and the mean employment share in manufacturing are about 0.2. Hence (20) implies that the elasticity, when evaluated at the mean and taking $\rho_{c,j}^{M}$ to be parametric, is roughly equal to $\rho_{c,R}^{M} - \rho_{c,N}^{M}$, which according to the Table 5 is on the order of magnitude of 0.1. Columns three to six contain the results for job-specific employment shares. It is clearly seen that refugee-rich counties have higher employment shares in job, which rely on learning-by-doing, are characterized by unskilled occupations, employ unskilled workers and rely on unskilled-biased technologies. Together with the findings on individual earnings, this is consistent with refugee inflows having triggered labor supply shifts towards unskill-intensive occupations and subsequently having induced demand shifts in precisely those industries.

To summarize, the consequences of this large scale population transfer are as follows. For the population of refugees, the expulsion was associated with substantial breaks in their employment histories that for many persisted through their entire career. At the sectoral level, there is a substantial increase in manufacturing employment both within individual's employment life-cycle and across cohorts. At the occupational level, the initial supply shift is associated with occupational downgrading in that we see large increases in the share of unskilled workers in the refugee population. I interpret these occupational choices as being indicative of refugees having a comparative advantage in industries, which rely relatively little on job-specific experience and human capital. This reallocation had sizeable effect on wages

	Empl. share in manufacturing s^{Man}	$ln(s^{Man})$	Job-specific employment shares			
share of refugees (s^{Ref})	0.928** (0.427)		-0.0199 (0.0130)	-0.0164 (0.0101)	-0.0643*** (0.0129)	-0.0443*** (0.0115)
log share of refugees ($ln(s^{Ref})$)		0.247** (0.118)				
$s^{Ref} \times$ Learning On Job			0.0278* (0.0162)			
$s^{Ref} \times$ Unskilled occup.				0.0383*** (0.0111)		
$s^{Ref} \times$ Low educational attain.					0.0813*** (0.0113)	
$s^{Ref} \times$ USB technology						0.0750*** (0.0113)
log population density (1975)	0.00601 (0.0195)	0.00434 (0.0399)	0.000109 (0.000443)	0.000111 (0.000443)	0.000106 (0.000443)	0.0000739 (0.000444)
empl share agricult. (1939)	0.313** (0.131)	0.669** (0.294)	0.00801*** (0.00281)	0.00803*** (0.00281)	0.00755*** (0.00280)	0.00834*** (0.00279)
empl share industry (1939)	0.811*** (0.153)	1.708*** (0.378)	0.00301 (0.00293)	0.00306 (0.00293)	0.00278 (0.00293)	0.00338 (0.00290)
Observations	245	245	12991	12991	12991	12991
R^2	0.278	0.293	0.004	0.003	0.006	0.004

Notes: Robust standard errors are shown in parentheses. *, ** and * denotes significance at the 1%, 5% and 10% level respectively. All regressions include a set of state fixed effects. The excluded instrument is "log population density in 1939" at the county level. In the first stage regression, "log population density in 1939" has a t-statistic of -7.58. The interaction terms are instrumented with the respective interactions with the log of the population density in 1939.**

Table 12: The effect of refugees on county level employment shares in 1975

in that even 15 years after the expulsion, refugees earn on average 10% less. About 70% of these lower wages can be explained by this adverse sorting of refugees into low-wage sectors and occupations.

At the level of local labor markets, the long-run effects of refugee inflows are positive as average wages (conditional on individual characteristics) are higher in refugee-rich counties. The instrumental variable estimate predicts that a 10% higher refugee share in 1950 increases wages by about 5% in 1975. This positive effect on wages is especially large for unskilled workers and younger cohorts and in jobs, relying on technologies which are biased towards unskilled workers. Additionally, the initial allocation is a strong predictor of the size of the manufacturing sector and employment share in unskill-biased occupations. These results are consistent with the theoretical framework and suggests biased technological adoption, which is sufficiently strong for relative labor demand being upwards sloping. While the data allows me to measure job-specific technologies, I am not able to distinguish between the extensive and the intensive margin of this adoption, i.e. did refugee-rich counties experience entry of unskill-intensive firms or takes adoption place within the firm. In the theoretical framework, I used entry as a tractable way to generate endogenous bias. Incidentally, there is anecdotal evidence for this modelling device. In 1949, Clarence M. Bold, the deputy director of the regional office for the military government in Bavaria, argued that "since refugees and bombed-out Bavarians now living in rural areas cannot move nearer to industrial jobs, such jobs must go to them. In fact many world famous industries wanting to reestablish in Bavaria have already sought locations in non-industrial areas near idle workers" (Office of the Military Government for Germany, 1949, p. 26). By using data on firms' location decision in the 50s and 60s, it would be possible to study if this was indeed an important margin of adjustment and I plan to do so in the near future.

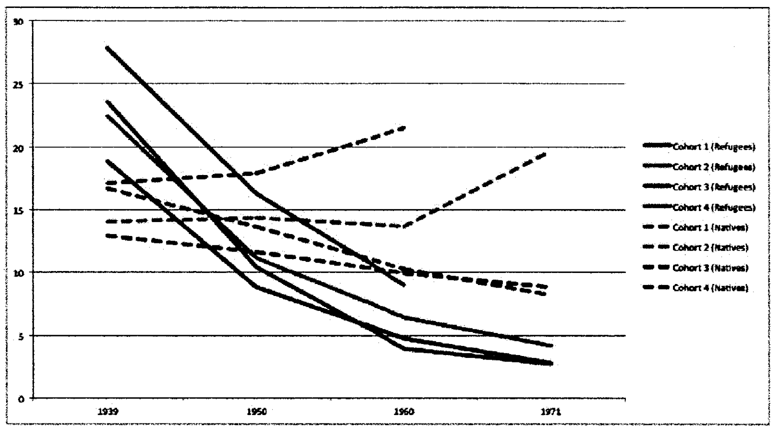
4 Conclusion

If firms change their technology in response to changes in labor supply, supply shifts will induce movements in the long-run aggregate demand function. If technological adoption is biased towards the more abundant factor of production, the long-run demand function will be more elastic and might be upward sloping if the technological response is strong enough. To know how firms' production possibility sets react to changes in labor supply is therefore crucial to predict the long-run implications of supply shifts on factor prices and allocation patterns. In this paper I study this question by exploiting a historical episode that induced large shifts in labor supply in Western Germany. After World War II, 8 million ethnic Germans were expelled from their homes in Eastern Europe and transferred to Western Germany. This increased the population by 20% in a 2-year period. Analyzing individual-level data from the 1960s and 70s, I show that the population of refugees experienced drastic sectoral and occupation change. In particular, the share of manufacturing workers increased by 45% and the employment share in unskilled occupations by 55% at the impact of the expulsion and this reallocation persisted through the entire employment life-cycle of most refugees. Refugee inflows can therefore be interpreted as a shift in labor supply, which is particularly pronounced in unskill-intensive occupations. To test if firms endogenously adopted biased technologies as a response, I exploit the regional variation in refugee inflows across local labor markets. Using information on detailed job-characteristics, I show that 25 years after the initial allocation, refugee-rich counties have a larger manufacturing sector and are characterized by larger employment shares in unskill-intensive occupations and in jobs where formal human capital is unimportant and unskill-biased technologies predominantly used. Turning to the wage data, I find that wages in refugee-rich counties are not only higher on average but especially so in precisely these occupations. This is consistent with theories of biased technological adoption, where the inflow of refugees induced shifts in the demand function for inexperienced labor.

5 Appendix

5.1 Cohort profiles for agricultural employment

Figure 5 shows the cohort-specific employment share in agriculture. The increase in the agricultural share at the late stages of the life-cycle is again likely due to earlier retirement in non-agricultural sectors (see also footnote 19).



Notes: The figure shows the employment shares in agriculture for different cohorts in both the refugee and the native population. The results for refugees are shown in solid lines, the results for natives are the dashed lines.

Figure 5: Employment share in agriculture for different cohorts

5.2 BBIB Data

Table 13 presents the results of probit estimates of the probability of using unskilled biased technologies (columns 1 to 3) and of having learnt the required skills on the job (columns 4 to 6).

	Usage of unskill-biased machines			Skills learnt on the job		
High education	-0.648***	-0.603***		-0.384***	-0.364***	
	(0.0236)	(0.0260)		(0.0215)	(0.0237)	
log income		-0.470***			-0.0343	
		(0.0318)			(0.0235)	
Firm has less than 10 employees			0.0187			0.0209
			(0.0352)			(0.0311)
Firm has more than 1000 employees			-0.0453			-0.0366
			(0.0333)			(0.0309)
Observations	16820	14495	16820	16820	14495	16820

Notes: Robust standard errors shown in parentheses. ***, ** and * denotes significance at the 1%, 5% and 10% level respectively. All regressions include state fixed effects, sector fixed effects and fixed effects for 7 categories of city size.

Table 13: The correlates of unskilled-biased technologies and learning on the job

Chapter 3

Firm Heterogeneity and Import Behavior: Evidence from French Firms[1]

1 Introduction

Trade in intermediate inputs is an important component of aggregate trade flows and has been growing at an enormous pace in the last decades (Feenstra, 1998; Hummels, Ishii, and Yi, 2001). Furthermore, a number of recent papers have established that trade in intermediate inputs can have substantial positive effects on firm productivity (Amiti and Konings, 2007; Goldberg, Khandelwal, Pavcnik, and Topalova, 2009; Halpern, Koren, and Szeidl, 2009). An analysis of firms' import demand is therefore central for our understanding of both aggregate trade flows and firm level productivity gains from international sourcing. So far, however, relatively little is known about firms' import behavior.[2] In particular, a theoretical framework of importing that is consistent with the firm-level evidence is missing. In this paper we take a step towards filling this gap. We start by proposing a simple model of import demand which is based on two core assumptions: productivity differences are the only source of firm-level heterogeneity and prices, fixed costs and input qualities are common across firms. We then use a comprehensive data set of French manufacturing firms to test its qualitative predictions on both the extensive and intensive margin. In this way, we identify the aspects of the data which such simple model can account for, and the ones for which further theoretical refinements are needed. Furthermore, we show the precise direction in which the core assumptions of the model need to be amended to qualitatively match the micro data.

In the last decade, the field of international trade has seen a fruitful interaction between theoretical work and micro evidence. However, the vast majority of these contributions has been concerned with the behavior of exporters. A first generation of empirical papers established a number of stylized facts, namely that few firms engage in exporting, that exporters are larger and more productive than non-exporters, and that exporters usually sell most of their output domestically (Bernard and Jensen, 1999; Bernard, Jensen, Redding, and Schott,

[1]This chapter is the product of joint work with Joaquin Blaum and Claire Lelarge. We especially thank Arnaud Costinot for numerous helpful discussions and persistent encouragement. We also thank Daron Acemoglu, Thomas Chaney, Sam Kortum and seminar participants at the University of Chicago and MIT.

[2]Notable exceptions are Bernard, Jensen, and Schott (2009); Bernard, Jensen, Redding, and Schott (2007), Halpern, Koren, and Szeidl (2009) and Kugler and Verhoogen (forthcoming, 2010).

2007). These findings were accompanied, on the theory side, by the development of frameworks that allow for firm-level heterogeneity and assume fixed costs to exporting (Melitz, 2003; Bernard, Eaton, Jensen, and Kortum, 2003; Chaney, 2008). More recently, Eaton, Kortum, and Kramarz (2004, 2008) have uncovered a number of novel facts regarding the entry behavior of firms into different export markets, and have subsequently shown how these facts can be theoretically rationalized by augmenting Melitz (2003) with various dimensions of firm-level heterogeneity as well as the marketing cost function of Arkolakis (2008). Compared to the developments on the export side, our knowledge about firms' import decisions is rather limited. Bernard, Jensen, Redding, and Schott (2007) show that the basic firm level facts about exporters are replicated for importers in the US, namely that importing is a rare activity and that importers are bigger and more productive than non-importers[3]. More recently, Halpern, Koren, and Szeidl (2009) establish some new firm-level facts on entry into import markets, and estimate a structural model of import behavior to identify and quantify the mechanisms driving the productivity gains associated with imported inputs.[4]

To motivate the basic ingredients of the theory, we start by documenting a set of facts regarding the entry behavior of importers into different import markets. First, we find substantial heterogeneity in the number of imported products across firms: while the median importing firm sources only 6 products, many firms source more than 40. Second, we study a new entry margin by defining an import market as an 8-digit (NC8) product from a given country, that is, as a variety.[5] We document the distribution of the average number of varieties per product French importers source, and find that the vast majority sources a limited number of varieties - 90% of the firms use less than three varieties per product. These facts are suggestive of firm-level differences in productivity, and non-trivial fixed costs to importing additional varieties. Finally, on the intensive margin, we document a striking degree of concentration in that importers spend most of their import expenditure on a small number of products and varieties: even intense importers, sourcing more than 50 products or 25 varieties per product, spend 80% of their expenditure on the top 5 products or varieties respectively. This suggests that quality differences across varieties are an important aspect of the data.

Based on these findings, we build a theoretical framework featuring firm-level productivity differences and variety-specific fixed costs and qualities. Firms are monopolistic producers of a differentiated good (as in Melitz (2003)) and decide on the cost-minimizing way to produce.

[3]In fact, this is reasonable given that importing and exporting are highly correlated.

[4]While Halpern, Koren, and Szeidl (2009) is closely related to this paper in that they also study firms' import demand, their emphasis is very different. Their main contribution is to quantify the importance of the quality and the complementarity channels by which imported inputs increase firm productivity. In contrast, we are interested in a model that can successfully account for the intensive and extensive margins of import demand.

[5]This definition of a variety is standard in the literature. See Broda, Greenfield, and Weinstein (2006).

There is a set of potential inputs, each of which is available in a domestic variety and possibly many foreign varieties. The production technology features a love-of-variety effect, but the firm's ability to reap these benefits is limited by the presence of two types of fixed costs - to being an importer and to importing any specific variety. The model is standard in the following sense: (i) there is only one dimension of firm level heterogeneity (namely, productivity), and (ii) input prices, trade costs (both fixed and variable) and the qualities of imported intermediates are not allowed to be a function of firms' characteristics[6], though they can differ across varieties.

On the extensive margin, the model predicts that more productive firms select into importing, import more products, and import more varieties of each product. The reason is that more productive firms benefit more from a given reduction in the marginal cost of production and are therefore willing to pay more for the possibility of using an extra input, or an extra variety of a given input (Melitz, Helpman, and Yeaple, 2004). On the intensive margin, there are two main implications. The first implication of the model is intuitive: as more productive firms source more products (varieties), the model predicts that these firms spend a lower share of their expenditure on *each* product (variety)[7]. The second implication is more subtle. Our assumptions of common input prices and qualities across firms and factor neutral productivity imply that conditional on the sourcing strategy (i.e. which products and varieties to import), expenditure shares across products and varieties are fully determined by price-adjusted qualities. Firm-characteristics, in particular productivity, should not affect firms' relative input demand once the sourcing strategy is controlled for.

We then test these predictions using a dataset of French manufacturing firms. Overall, the model does fairly well when contrasted to the data. Concerning the extensive margin predictions, we find that more productive firms are more likely to be importers, import more products and more varieties per product. On the intensive margin, we first analyze the allocation of expenditure across imported inputs. We find that more productive firms have indeed lower expenditure shares for their individual varieties. More importantly, once we control for firms' sourcing strategies[8], firm-productivity and other firm-characteristics cease to be important determinants. That is, we find that productivity affects foreign expenditure shares only through its effect on the sourcing strategy. We regard this as evidence supporting the two core

[6]This contrasts with the assumptions made in the export literature (Eaton, Kortum, and Kramarz, 2008; Arkolakis, 2008).

[7]While this prediction seems to be almost mechanical, it is important to note that it would not hold in a variety of alternative settings. If, for instance, productivity was factor biased (see Kugler and Verhoogen (2010) for an example), an increase in productivity may increase the number of products sourced but some products would nonetheless receive a higher expenditure share. Thus, testing the prediction that more productive firms devote a lower share to *each* of their products is informative about the underlying production process.

[8]Recall that the sourcing strategy is defined as the set of products and varieties imported.

ingredients of our standard model (namely factor neutral productivity and common prices and qualities across firms) when applied to the relative demand across foreign inputs. However, when we turn to firms' expenditure on domestic vs. foreign inputs, the model is not supported by the data. We show that after controlling for the sourcing strategy firm characteristics significantly affect domestic spending decisions. In particular, for a given sourcing strategy, more productive firms spend a higher share of their budget on domestic inputs. If French products are of high quality, this result is consistent with firm productivity and input quality being complements (see also Kugler and Verhoogen (2010)).

Taken together, these results indicate that a standard model of import demand with a single source of firm level heterogeneity and prices that cannot vary with firm characteristics is qualitatively consistent with the extensive margin evidence but is at odds with an important aspect of the intensive margin. While these assumptions seem to work reasonably well for spending decisions across foreign varieties, they are too restrictive to account for the pattern of expenditure on domestic inputs. In particular, a mechanism inducing firm level variation in the effective relative price of domestic inputs vis-a-vis foreign varieties (above and beyond the variation induced by the sourcing strategy) is required by the data. While an exact microfoundation for such mechanism is beyond the scope of this paper, our results highlight the precise dimension in which the standard model needs to be augmented.

The rest of the paper is organized as follows. Section 2 documents some extensive and intensive margin facts about importers. Section 3 contains our theoretical framework to study the behavior of importers. Section 4 brings the predictions of the model to the data and discusses the implications of the results. Finally, section 5 concludes.

2 Some Stylized Facts of Import Behavior

In this section we establish a number of novel facts about the import behavior of manufacturing firms. To this end, we use a comprehensive dataset containing all international transactions, as well as other important characteristics, of every manufacturing firm in the French economy for the years between 2001 and 2006. The main picture that emerges is one of high concentration among many different dimensions of the data. Few firms import but those who do spend most of their expenditure domestically. Regarding the import budget, even intense importers allocate most of it to a limited number of products and varieties.

2.1 Data

In this section we briefly describe our dataset. A more detailed description is contained in the Appendix. Because we are interested in the demand for inputs, we restrict the analysis

	Quantiles				
	25%	50%	75%	90%	99%
No of annual trade interactions per country	10	53	328	3,763	69,371
No of annual trade interactions per product	8	25	74	179	752
No of annual trade interactions per firm	2	8	25	59	238
Value of firm-product-country interactions (1000EUR)	44.15	268.53	1,258.83	5,319.10	70,993.13
Number of firm-product-country observations			705,316		

Table 1: The concentration of French imports.

to manufacturing firms. We observe import flows for every manufacturing firm in France from the official custom files. Manufacturing firms account for 31% of the population of French importing firms and 56% of total import value in 2001. Overall, French firms trade with a total 226 countries. The flows are classified at the 8-digit (NC8) level of aggregation, which means that the product space consists of roughly 9,500 products. Using unique firm identifiers we can match this dataset to fiscal files, which contain detailed information on firm characteristics. The final sample consists of an unbalanced panel of roughly 260,000 firms which are active between 2001 and 2006[9].

A first overview over our dataset is contained in Table 1, which shows characteristics of the distribution of the number of annual variety flows (i.e. product-country cells) across firms.[10] In total we observe roughly 700,000 variety-firm pairs. Given that there are about 30,000 importers in our data, the average importer imports about 23 varieties of potentially different products. Table 1 however shows that this average is not too informative as international activity is highly concentrated. The first two rows characterize the distribution of shipments across countries (row one) and products (row two) and show that French import flows are very concentrated - both geographically as well as in the product space. The median country is only active in 50 firm-product cells, whereas the top two exporting countries to France, namely Germany and Italy, report 70,000 interactions in distinct firm-product cells. A similar picture emerges when we look at the different products the French economy sources from abroad. For half of the potential products, i.e. roughly 5,000 products, only 25 country-firm interactions are observed, whereas the most popular products are shipped into France in more than 750 distinct country-firm combinations. Hence, international trade is very concentrated on few countries and few products. Note in particular the still very large difference between the 99% and 90%-quantile.

[9]This dataset is not new and has been used in the literature before (e.g. Mayer, Melitz, and Ottaviano (2010); Eaton, Kortum, and Kramarz (2004, 2008)). However, these contributions focused almost exclusively on the export side.

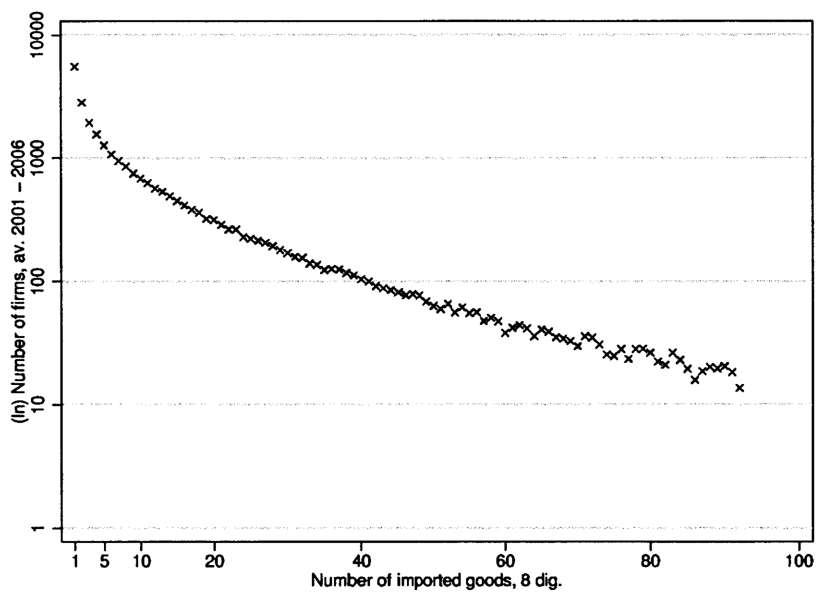
[10]Table 1 is of a similar flavor as the discussion of sparsity by Armenter and Koren (2008) but it is slightly different. Whereas they analyze the data on the *flow* level, we aggregate the data within a *firm-variety cell*, because it is these quantities which our model can speak to.

A similar concentration is also seen on the firm level. For most of the importing firms receiving an imported product is a rare event: the median firm sources only 8 varieties a year internationally. The top one percent of firms (that is, 300 firms) import 240 varieties. Finally, there is also a striking degree of heterogeneity in the total values different firms import from individual varieties. Whereas the most active firm-variety pairs are worth more than 70m EUR, a quarter of French importers import less than 45,000EUR worth of the varieties they source internationally within a year. Thus, at whatever dimension we look, the world of international trade is a small world. Few firms are actively participating and, when they do, they tend to import from a small set of countries and only a small subset of potential commodities.

2.2 Facts on Import Behavior

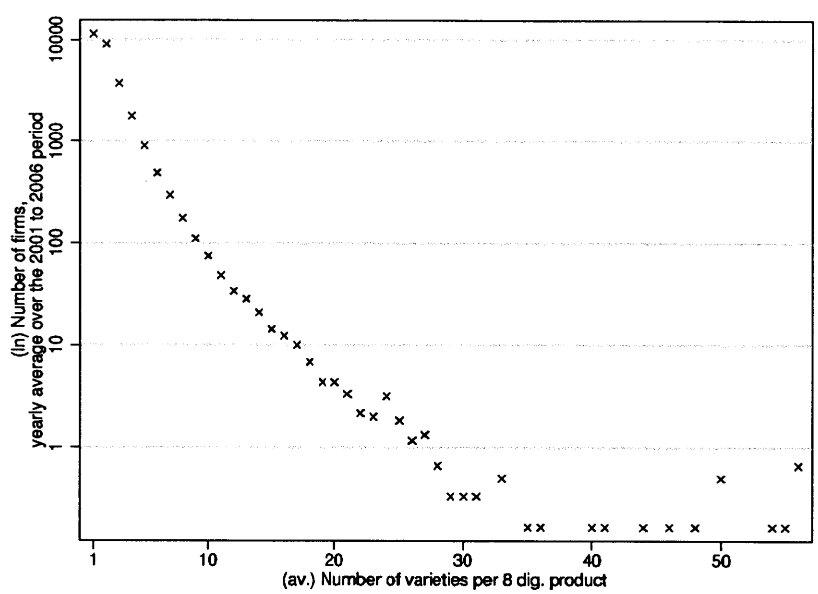
Fact 1. There is substantial heterogeneity in the number of products imported. The median importer sources 6 products from abroad, but some firms source up to 90 products. Figure 1 shows the number of firms importing k products for different values of k. Three properties stand out. First of all, there is substantial variation in the number of products sourced across importing firms. Whereas roughly 20% of firms source only one product, there are a number of firms importing more than 40 products from abroad. On average, French importers import 13 products from abroad, but due to the skewness of the distribution the median number of products is only 6. Secondly, the locus depicted in Figure 1 is strictly decreasing and for the more intense importers (i.e. firms importing more than 10 products), the graph is almost linear, so that the semi-elasticity of the "product extensive margin" is roughly constant. Increasing the number of imported products by one decreases the number of firms doing so by 1.4 percent. Finally, the schedule is highly convex, i.e. there are *much* fewer firms sourcing exactly 10 products than there are firms sourcing only one product.

Fact 2. Most firms source few varieties of each product. 90% of the firms source at most 3 varieties per product, and only 1% sources more than 8. In Figure 2 we depict the distribution of the average number of varieties per product. Figure 2 clearly shows that the vast majority of firms sources only a very limited number of varieties - 90% of French firms have less than three varieties per product and only one percent of the firms source more than 8 varieties per product. So if there are indeed high productivity gains through love of variety effects (Halpern, Koren, and Szeidl, 2009; Goldberg, Khandelwal, Pavcnik, and Topalova, 2009), global efficiency gains from a finer division of labor are far from fully exploited.



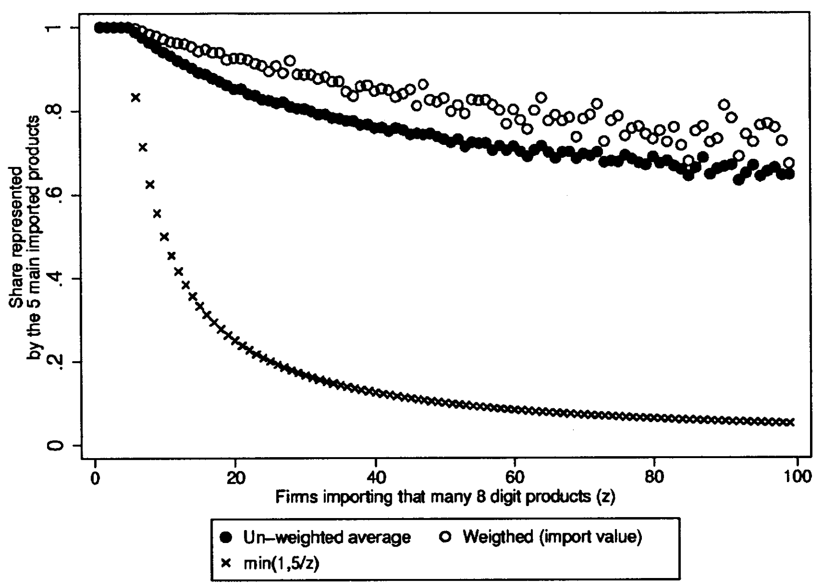
Notes: The figure shows the number of firms importing k products for different numbers of k. A product is defined on the 8-digit level and a firm is defined to be an importer of product k, whenever it imports a positive amount from at least one country. We use 6 years of data from 2001-2006 and report the yearly average.

Figure 1: How many products do firms import?



Notes: The figure shows the distribution of the average number of varieties per product, i.e. the distribution of $\overline{V}_i = \sum_k s_{k,i} V_{k,i}$, where $s_{k,i}$ is the expenditure share of firm i on product k and $V_{k,i}$ is the number of varieties firm i sources in product k. We use 6 years of data from 2001-2006 and report the yearly average.

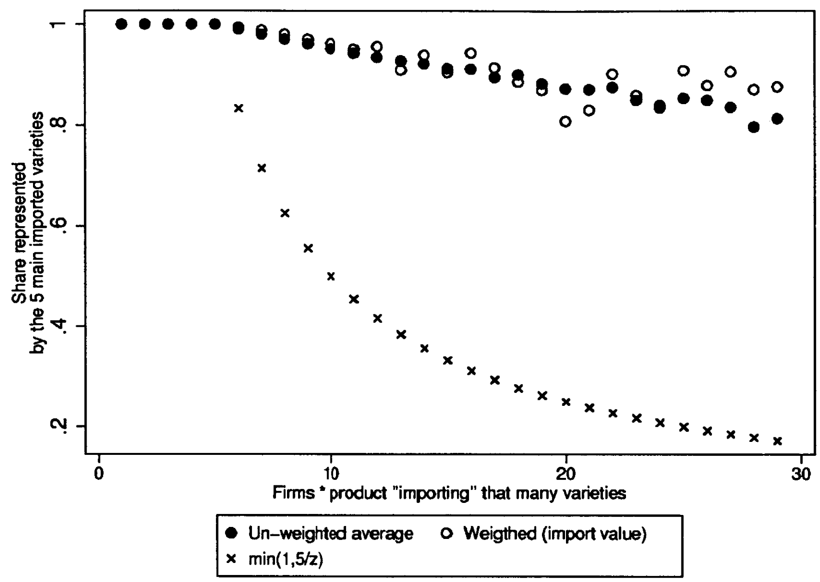
Figure 2: How many varieties do firms source?



Notes: The figure shows the average expenditure share on the 5 most popular products of firms importing k products for different values of k. For the unweighted average all firms get an equal share, for the weighted results, we weigh firms by their import value. We also depict the counterfactual expenditure share if expenditures were equalized across products.

Figure 3: Concentration of firms' import spending across products.

Fact 3. Expenditure shares are highly concentrated across products. Firms importing 50 products spend 80% of their import budget on only 5 products. Importers spend an overwhelming part of their import expenditure on a small number of products. In Figure 3 we show the cumulative average expenditure share of the five most popular products for firms sourcing k products for different values of k. Given that firms who source only 6 or 7 products spend *by construction* at least a fraction of 72%-83% on their five most popular products, we also depict the counterfactual expenditure share if expenditures were equalized across products. It is clearly seen that the concentration is high. Firms sourcing 50 product still allocate 80% of their expenses on merely 5 products and even firms that source 100 products do so with 70% of their budget.



Notes: The figure shows the average expenditure share on the 5 most popular varieties of firm-product pairs with V many varieties for different values of V. For the unweighted average all firms get an equal share, for the weighted results, we weigh firms by their import value. We also depict the counterfactual expenditure share if expenditures were equalized across varieties.

Figure 4: Concentration of firms' import spending across varieties.

Fact 4. Expenditure shares are highly concentrated across varieties within products. Firms importing 25 varieties per product spend 80% of the product's expenditure on only 5 varieties. We now look at the allocation of expenditure within products. In Figure 4 we depict the average expenditure share on the top 5 varieties for firms sourcing V varieties per product for different values of V. Even firms sourcing 25 varieties, spend 80% of their total expenditure on those varieties on their 5 most popular ones.

3 A Simple Model of Importing

Motivated by these facts, we develop a theoretical framework to account for firms' import decisions. First, we allow for productivity differences across firms to be able to account for the startling degree of heterogeneity in firms' behavior. Second, we allow for variety-specific fixed costs of importing to rationalize the fact that most firms source a rather small number of products and varieties from abroad (Facts 1 and 2). Finally, we allow for differences in quality and transport costs across imported intermediates. This seems necessary in face of the high variation in expenditure shares documented in Facts 3 and 4. We then impose two further assumptions that will play a key role in the results. First, we allow for a single dimension of firm-level heterogeneity (factor neutral productivity). Second, we restrict prices, product qualities and trade costs (both fixed and variable) to be common across firms. These two core assumptions constitute a natural starting point, and give empirical content to the model.[11] We incorporate these elements in an otherwise standard model of international trade, which is probably the most basic model applied work would start with. This will allow us to identify the features of the data that such canonical model can and cannot explain, and thus to isolate avenues for future research.[12]

We consider the following economy. On the output side, we adopt a specification similar to Melitz (2003). There exists a continuum of firms, each of which is a monopolist in its particular product line. As stressed above, the single dimension of firm heterogeneity is their (factor neutral) productivity. Formally, each firm has access to a production function

$$y = \varphi q(z), \tag{1}$$

where φ denotes the firm's TFP, z is a vector of inputs and q is a constant returns to scale production function. To keep the model close to the empirical analysis, we make the usual distinction between products and varieties. In particular, we think of inputs as being different *products* when they come from different 8-digit product classes and differentiated *varieties* if they are in the same product class but sourced from different countries. Hence, $z = [z_{jk}]_{j,k}$, where k denotes a particular product and j denotes the variety of this product. For conve-

[11]These assumptions contrast with the ones made in Halpern, Koren, and Szeidl (2009), who assume fixed costs to vary at the firm level, and Kugler and Verhoogen (2010), whose model generates non-neutral productivity differences across firms.

[12]The recent years have seen substantial progress on the export side, with Eaton, Kortum, and Kramarz (2008) allowing for firm-specific demand shocks, Arkolakis (2008) developing a microfoundation for heterogeneous fixed costs of entry and ? considering multi-product firms. These are examples of theoretical advances being influenced by empirical failures of a standard model. As the theory of importing is less developed than its exporting counterpart, an empirical investigation of the qualitative patterns of a standard model will hopefully have similar consequences there.

nience[13] we model q as a nested CES production function

$$q(x_1, ..., x_K) = \left(\sum_{k \in K} x_k^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon-1}} \text{ where } x_k = \left(\sum_{j \in V_k} (\eta_{jk} z_{jk})^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}},$$

where η_{jk} measures the quality of variety j for product k and V_k denotes the set of varieties of product k.

On the demand side, consumers' preferences for the differentiated products are given by the usual CES aggregator

$$Y = \left(\int_0^1 y(\varphi)^{\frac{\sigma-1}{\sigma}} d\varphi \right)^{\frac{\sigma}{\sigma-1}},$$

where $y(\varphi)$ is the level of output of firm φ. Assuming that $\sigma > 1$, the demand for firm φ's product is given by

$$y(\varphi) = Y p(\varphi)^{-\sigma}, \tag{2}$$

so that the profit maximizing price is $p(\varphi) = \frac{\sigma}{\sigma-1} MC(\varphi)$, where $MC(\varphi)$ denotes the firm's marginal costs, and we use the ideal price index as the numeraire. Firms' marginal costs depend on their production decisions, in particular on their international sourcing strategy. We order varieties such that the first variety for each product refers to the domestic input. Firms take the set of prices $[p_{jk}]_{j,k}$ as given. These prices contain all variable transport costs which accrue whenever a foreign variety is acquired, that is, $p_{jk} = (1 + \tau_{jk}) p^*_{jk}$, where p^*_{jk} is the price of product k at country j and τ_{jk} are the iceberg transport costs of product k from country j to France.

Sourcing a foreign input is subject to fixed costs. In particular, sourcing variety j of product k from abroad entails a fixed cost of $f_{jk} > 0$. Additionally, there is a fixed cost to being an importer, which we denote by $\kappa > 0$. We assume that there are no fixed costs to sourcing domestic varieties. The firm's cost-minimization problem consists therefore in choosing a sourcing strategy $V = (V_1, .., V_K)$ - which specifies a set of inputs K and sets of varieties V_k for each input- as well as input demands z_{jk}.

Consider a firm that has decided to follow a sourcing strategy given by V. Using the CES

[13] All of the theoretical results do not rely on the CES structure. Imposing the CES structure, however, gives us convenient closed form solutions for the demand systems, which will guide our empirical implementation in Section 4.

structure, standard calculations show that unit costs are given by

$$c(V_1,...,V_K) = c(V) = \left(\sum_{k \in K} \left(\sum_{j \in V_k} (\xi_{jk})^{\rho-1} \right)^{\frac{\varepsilon-1}{\rho-1}} \right)^{\frac{1}{1-\varepsilon}}, \tag{3}$$

where

$$\xi_{jk} \equiv \frac{\eta_{jk}}{p_{jk}} = \frac{\eta_{jk}}{(1+\tau_{jk})\, p^*_{jk}} \tag{4}$$

is the quality flow per dollar spent of variety j in product k. Using (3), the marginal cost of production is therefore given by

$$MC(\varphi, V) = c(V) \frac{1}{\varphi}. \tag{5}$$

Note that $c(V)$, and hence $MC(\varphi, V)$, is strictly decreasing in the number of elements in V_k as the production function features a standard love of variety effect. This captures the efficiency gains from being active internationally.

Let us now turn to the choice of the optimal sourcing strategy. Using (5), the demand function (2) and the profit maximizing price, firms' profits are simply given by

$$\pi(V;\varphi) = \begin{cases} A\left(\frac{\varphi}{c(V)}\right)^{\sigma-1} & \text{if domestic} \\ A\left(\frac{\varphi}{c(V)}\right)^{\sigma-1} - \sum_{k\in K(\varphi)} \sum_{j\in V_k(\varphi)} f_{jk} - \kappa & \text{if importer} \end{cases} \tag{6}$$

where $A = \left(\frac{1}{\sigma-1}\right)^{1-\sigma} \left(\frac{1}{\sigma}\right)^{\sigma} Y$ is constant across firms. Given (6), the optimal sourcing strategy of firm φ is then given by

$$V(\varphi) = (V_1(\varphi), V_2(\varphi), ..., V_K(\varphi)) = \arg\max_V \pi(V;\varphi). \tag{7}$$

Given the complementarity between $c(V)$ and φ contained in (6), it is immediate that $c(V(\varphi))$ is weakly decreasing in φ so that more productive firms produce at lower marginal costs. In particular, there exists a critical level of productivity $\overline{\varphi} = \overline{\varphi}(f, \kappa, \xi)$ such that only firms with $\varphi > \overline{\varphi}$ are active internationally. Intuitively, more productive firms gain more from lowering their marginal costs and hence are more willing to pay the fixed costs to access a higher number of varieties (Melitz, Helpman, and Yeaple, 2004).

This heterogeneity in sourcing strategies has also implications for the distribution of sales.

In particular, sales are given by

$$ln\left(p\left(\varphi\right)y\left(\varphi\right)\right) = ln\left(\sigma A\right) + \left(\sigma - 1\right)\left(ln\left(\varphi\right) - ln\left(c\left(V\left(\varphi\right)\right)\right)\right), \tag{8}$$

which now depend on firm productivity via the direct productivity effect ($ln\left(\varphi\right)$) and indirectly through the sourcing strategy ($V\left(\varphi\right)$). As $c\left(V\left(\varphi\right)\right)$ is decreasing in φ, sales are not only strictly increasing in productivity, but the elasticity of sales with respect to productivity exceeds $\sigma - 1$, the one of the closed economy.

Let us now turn to the set of implications which are more specific to firms' import behavior. First, we consider the number of varieties sourced internationally. While (7) shows that the marginal cost of production $c\left(V\left(\varphi\right)\right)$ is decreasing in productivity via the endogenous sourcing decision, this does *not* imply that productive firms source *more* varieties. It might be that unproductive firms source multiple varieties with low fixed costs and low quality flows and high productivity firms concentrate on few fixed cost expensive varieties which yield high quality flows ξ. We can show, however, that under a mild condition[14] the model predicts that less productive firms source a strict subset of the varieties their more productive counterparts import. This will be assumed from now on[15]. As a corollary, the model predicts that the number of importer products,

$$K\left(\varphi\right) = \sum_{k \in K} 1\left[V_k\left(\varphi\right) > 1\right],$$

is increasing in firm productivity.

The model also has strong implications for the intensive margin of trade. Using the CES structure, firm φ's expenditure share on variety v for product k is given by

$$s_{vk}\left(\varphi\right) = \frac{p_{vk}z_{vk}}{\sum_{j \in V_k(\varphi)} p_{jk}z_{jk}} = \frac{\left(\eta_{vk}/p_{vk}\right)^{\rho-1}}{\sum_{j \in V_k(\varphi)} \left(\eta_{vk}/p_{vk}\right)^{\rho-1}} = \frac{\xi_{vk}^{\rho-1}}{\sum_{j \in V_k(\varphi)} \xi_{jk}^{\rho-1}} = \frac{1}{V_k\left(\varphi\right)}\left(\frac{\xi_{vk}}{\xi_k\left(\varphi\right)}\right)^{\rho-1} \tag{9}$$

where $\xi_k\left(\varphi\right) = \left(\frac{1}{V_k(\varphi)} \sum_{j \in V_k(\varphi)} \xi_{jk}^{\rho-1}\right)^{\frac{1}{\rho-1}}$ is the average quality flow per variety of product k. (9) has two main implications which will be at the center of the empirical analysis. First, more productive firms are predicted to have lower expenditure share on *each* individual variety.[16] Second, given a sourcing strategy $V\left(\varphi\right)$, expenditure shares are entirely determined from the distribution of price-adjusted qualities across varieties. It is important to note that our

[14]We show in the appendix that to a first order approximation this will be satisfied whenever the relative efficiency gains of an additional variety for two firms do not exceed their relative sales.

[15]With a slight abuse of notation, $V_k\left(\varphi\right)$ will denote both the set of varieties and the total number of varieties that firm φ sources of input k.

[16]Recall that this implication is not mechanical. See the discussion in footnote 7.

two core assumptions (namely, factor neutral productivity as the single source of firm level heterogeneity and prices, fixed costs and qualities being common across firms) imply that expenditure shares *only* depend on φ via the set of varieties sourced. This is an exclusion restriction which we will test in the data. Also note that this restriction does not rely on the CES functional form.

Since expression (9) also holds for the domestic variety, the model predicts that more productive firms spend relatively less on domestic inputs. However, for a given choice of imported varieties, domestic expenditure shares should not depend on productivity. The firm's total expenditure share on domestic products $s^D(\varphi)$ is a weighted average of product-specific domestic shares $s_k^D(\varphi)$, i.e.

$$s^D(\varphi) = \sum_{k\in K} s_k(\varphi)\, s_k^D(\varphi) = \sum_{k\in K} \frac{\left(\sum_{j\in V_k(\varphi)} \xi_{jk}^{\rho-1}\right)^{\frac{\varepsilon-1}{\rho-1}}}{\sum_{k\in K}\left(\sum_{j\in V_k(\varphi)} \xi_{jk}^{\rho-1}\right)^{\frac{\varepsilon-1}{\rho-1}}} s_k^D(\varphi). \qquad (10)$$

Again, (10) shows that firms' domestic expenditure share depends on the price-adjusted qualities and on firms' sourcing strategies $V(\varphi)$ - conditional on the latter all firm characteristics are irrelevant.

To sum up, the main lessons we draw from this theoretical exercise are:

1. Importers are positively selected on productivity, i.e. the set of importers is given by $\mathcal{I} = \{\varphi | \varphi \geq \overline{\varphi}(f, \kappa, \xi)\}$, and sales are increasing in firm-productivity (see (8)).

2. The number of imported products is increasing in firm productivity φ.

3. Within products, the number of varieties firms import is increasing in firm productivity φ.

4. Expenditure shares across products and varieties are decreasing in firm-productivity. Furthermore, conditional on firms' sourcing strategies, expenditure shares across varieties and products are fully determined by the distribution of qualities $[\xi_{vk}]_{v,k}$ and thus do not depend on firm characteristics.

5. In particular, more productive firms have lower expenditure shares on domestic products.

In the following section we test these five predictions. Before doing so, we note that the model is consistent with the broad facts from the previous section. Dispersion in firm-level productivity translates into dispersion in the number of products sourced (Fact 1), and the number of varieties (Fact 2)[17]. Facts 3 and 4 can be rationalized by allowing the qualities and/or trade costs to have enough variation across products and varieties.

4 Testing the qualitative predictions of the model

The goal of this section is to test the five qualitative predictions derived above. An important issue that arises is that firm-level productivity is unobservable. We deal with this problem by taking the model at face value and proxying productivity with log sales, as (8) predicts a strictly monotone and roughly log-linear relation. As additional robustness checks we also consider two direct estimates of firm productivity. First we take an OLS-based measure by calculating TFP as the residual from industry-specific OLS regressions of log value-added on log capital and log employment (see Bernard and Jensen (1999); Kugler and Verhoogen (forthcoming)). Second we estimate firm-level TFP using the procedure proposed by Levinsohn and Petrin (2003).[18]

4.0.1 Prediction 1: Selection on productivity

The fundamental sorting prediction of the model is that importers are positively selected on productivity. Table 2 reports the results of a regression of an import dummy on measures of firm productivity and various other firm characteristics. The prediction is borne out in the data. Column (1) shows that more productive firms, as measured by sales, are more likely to be importers. In columns (2) and (3) we control for additional firm characteristics which potentially affect import status and are correlated with productivity. That the lagged import status is highly significant suggests that fixed costs to importing are at least partly sunk. It is reassuring that exporters and members of a foreign or corporate group are more likely to be importers. Note that column (3) contains firm fixed effects. Finally, columns (4) and (5) show that these results are robust to using other measures of productivity. These findings are consistent with the literature (Bernard, Jensen, Redding, and Schott, 2007; Halpern, Koren, and Szeidl, 2009; Kugler and Verhoogen, forthcoming).

4.0.2 Prediction 2: More productive firms import more products

Another extensive margin prediction of the model is that more productive firms source more products. Table 3 reports the results of testing this basic implication of the model. Column

[17]Note that in the model's notation, Fact 2 depicts the histogram of $\overline{V}(\varphi) = \sum_{k \in K(\varphi)} s_k(\varphi) V_k(\varphi)$, where $s_k(\varphi)$ denotes the expenditure share firm φ spends on its imported product k.

[18]We rely on the specification based on value added, using intermediate inputs as an instrument. As a third measure of TFP we also consider the firm-level Solow residual ($TFP = Y/(L^{\alpha}K^{1-\alpha})$ where α is a sector-specific

Dep. var.	Indicator of import status mean = 0.196				
	(1)	(2)	(3)	(4)	(5)
log Sales	0.107***	0.028***	0.029***		
	(0.000)	(0.000)	(0.001)		
Lagged Import Status		0.755***	0.018***	0.019***	0.018***
		(0.001)	(0.004)	(0.004)	(0.004)
TFP-LP				0.009***	
				(0.001)	
TFP-OLS					0.010***
					(0.001)
Exporter			0.088***	0.085***	0.085***
			(0.002)	(0.002)	(0.002)
Corporate Group			0.013***	0.013***	0.012***
			(0.002)	(0.003)	(0.003)
Foreign Group			0.013***	0.013***	0.013***
			(0.005)	(0.005)	(0.005)
Firm FE	N	N	Y	Y	Y
R^2	0.396	0.750	0.875	0.878	0.879
Observations	1,107,962	858,456	858,394	637,780	639,016

Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels. All regressions include year fixed effect, age fixed effect and sector fixed effects. The dependent variable is an indicator of the firm's import status. "Lagged Import Status" is an indicator for the firm's import status in the previous year. TFP-OLS is calculated as the residual from an industry-specific OLS regression of log value added on log capital and log employment. TFP-LP is calculated using the method of Levinsohn and Petrin (2003). A firm is member of a foreign group if at least one affiliate or the headquarter is located outside of France. A firm is member of a corporate group if it is controlled by another firm or it has control over at least one affiliate.

Table 2: Selection into importing

Dep. var.	ln nb. imported products [ln] mean = [1.948] 17.657				
	(1)	(2)	(3)	(4)	(5)
log Sales	0.493***	0.436***	0.342***		
	(0.001)	(0.002)	(0.007)		
Exporter		0.292***	0.114***	0.140***	0.140***
		(0.007)	(0.007)	(0.007)	(0.007)
Corporate Group		0.024***	0.045***	0.070***	0.071***
		(0.006)	(0.007)	(0.008)	(0.008)
Foreign Group		0.365***	0.025**	0.039***	0.035***
		(0.008)	(0.010)	(0.011)	(0.011)
TFP-LP				0.068***	
				(0.005)	
TFP-OLS					0.106***
					(0.006)
Firm FE	N	N	Y	Y	Y
R^2	0.449	0.463	0.918	0.914	0.915
Observations	167,733	167,711	167,711	156,506	158,460

Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels. All regressions include age, 4-digit industry and year fixed effects. The dependent variable is the log of the number of imported products of firm i, i.e. $ln\,(K_i)$, where $K_i = \sum_k 1\left[V_{k,i} > 1\right]$ and $V_{k,i}$ is firm i's number of varieties of product k. TFP-OLS is calculated as the residual from an industry-specific OLS regression of log value added on log capital and log employment. TFP-LP is calculated using the method of Levinsohn and Petrin (2003). A firm is member of a foreign group if at least one affiliate or the headquarter is located outside of France. A firm is member of a corporate group if it is controlled by another firm or it has control over at least one affiliate.

Table 3: Productivity and the number of imported products

(1) shows that sales are strongly positively correlated with the number of imported products. Adding other firm level controls such as dummies for exporting status or membership in a foreign or corporate group reduces the coefficient only slightly and even after incorporating firm fixed effects in column (3) sales still have a very high positive effect.[19] As expected, exporters and members of corporate and foreign groups source more products internationally. Quantitatively, a 10% increase in sales is associated with a 3.5% increase in the number of products sourced. These results are robust to using other measures of productivity[20] (see columns (4) and (5)).[21] Taken together, Figure 1 (from Section 2.2) and Table 3 are in line with the type of sorting implied by the model.

4.0.3 Prediction 3: More productive firms import a higher number of varieties

The final extensive margin prediction of the model is that more productive firms import each product from more countries. We test this implication by regressing the average number of varieties firms source on measures of firm-level productivity and additional controls. The results are contained in Table 4. It is clearly seen that the coefficient on sales is positive and highly significant. In specification (2) we control for a host of firm-level characteristics which are likely to be important determinants of the demand for imported varieties. The coefficient on sales hardly changes. In column (3) we introduce firm fixed effects. The coefficient on sales decreases by a third but remains robustly positive. Furthermore it is seen that exporters and firms with foreign affiliates source more varieties per product. Somewhat surprisingly membership of a corporate groups is associated with sourcing less varieties although this effect vanishes once we include firm fixed effects.[22] Quantitatively, 10% larger sales are associated with an increase of the average number of varieties sourced by about 1%. Finally these results are qualitatively confirmed with other measures of productivity (TFP-OLS and TPF-LP reported in columns (4) and (5)) and hold in specifications at the firm-product level, that is when taking the number of varieties per product ($V_{k,i}$) as the dependent variable.[23] The

wage share). Since the results are similar with this third measure, we only report the first two.

[19] Note that the additional firm-level variables control for intra-firm trade, that is for trade between members of a corporate group in different countries.

[20] Note that TFP is itself a function of the number of products and varieties sourced, as TFP, when measured from firm revenues, is given by $TFP(\varphi) = \frac{\varphi}{c(V(\varphi))}$. However, $c(V(\varphi))$ is decreasing in φ if and only if more productive firms source more products, so that the results in Table 3 are still informative about the selection mechanism of the model.

[21] We do not report these TFP measures in specifications without firm fixed effects, because in the cross-section, they are uncorrelated with sales and negatively correlated with importing status, which makes them unsuitable in the context of a Melitz (2003) model.

[22] One possibility why members of a corporate group source less foreign varieties might be that manufacturing units within a group rely on a French trading partner within the group to obtain their foreign inputs.

[23] These results are not reported but available upon request.

Dep. var.	ln av. nb. varieties [ln] mean = [0,487] 1,998				
	(1)	(2)	(3)	(4)	(5)
log Sales	0.169***	0.151***	0.110***		
	(0.001)	(0.001)	(0.003)		
Exporter		0.090***	0.025***	0.033***	0.033***
		(0.003)	(0.003)	(0.004)	(0.004)
Corporate Group		-0.016***	0.009**	0.016***	0.017***
		(0.003)	(0.004)	(0.004)	(0.004)
Foreign Group		0.149***	0.011	0.014*	0.011***
		(0.004)	(0.007)	(0.007)	(0.007)
TFP-LP				0.023***	
				(0.002)	
TFP-OLS					0.035***
					(0.003)
Firm FE	N	N	Y	Y	Y
R^2	0.294	0.304	0.842	0.840	0.841
Observations	167,733	167,711	167,711	156,506	158,460

Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels. All regressions include age, 4-digit industry and year fixed effects. The dependent variable is the log of the average number of varieties of firm i, i.e. $ln(V_i)$, where $V_i = \sum_k s_{k,i} V_{k,i}$ and $V_{k,i}$ is firm i's number of varieties of product k and $s_{k,i}$ is firm i's expenditure share on product k. TFP-OLS is calculated as the residual from an industry-specific OLS regression of log value added on log capital and log employment. TFP-LP is calculated using the method of Levinsohn and Petrin (2003). A firm is member of a foreign group if at least one affiliate or the headquarter is located outside of France. A firm is member of a corporate group if it is controlled by another firm or it has control over at least one affiliate.

Table 4: Productivity and the Number of Varieties

results in Table 4 are therefore consistent with the sorting mechanism which lies at the heart of our model.

4.0.4 Prediction 4: Expenditure shares are decreasing in productivity. However, conditional on the sourcing strategy, productivity is irrelevant.

Consider now the intensive margin, that is the expenditure shares across differentiated varieties within imported products[24]. From (9) we have that

$$ln\left(s_{vk}\left(\varphi\right)\right) \quad = \quad -ln\left(V_k\left(\varphi\right)\right)+\left(\rho-1\right)\left(ln\left(\xi_{vk}\right)-ln\left(\xi_k\right)\right), \tag{11}$$

where $\xi_k = \left(V_k\left(\varphi\right)^{-1}\sum_j^{V_k(\varphi)}\xi_{jk}^{\rho-1}\right)^{\frac{1}{\rho-1}}$ is the average quality of the $V_k\left(\varphi\right)$ varieties of product k the firm sources. (11) shows that the expenditure share on individual varieties within products is decreasing in the number of varieties sourced of that product ($V_k\left(\varphi\right)$), and that it is decreasing in productivity (as $V_k\left(\varphi\right)$ is increasing in φ). Most importantly, conditional on $V_k\left(\varphi\right)$, firm characteristics, especially productivity, are excluded from the determination of expenditure shares. Intuitively, once the sourcing strategy is controlled for (which under Equation (16) reduces to the number of varieties), expenditure shares are only a function of price-adjusted qualities[25] which are assumed to be invariant across firms.[26] (11) suggest the following regression for each product-variety pair

$$ln\left(s_{vk}^i\right) \quad = \quad \alpha_{vk}-\gamma_{vk}ln\left(V_k^i\right)+\beta_{vk}ln\left(S^i\right)+\mu_{vk}X^i+u^i, \tag{12}$$

where all parameters are variety-product specific, s_{vk}^i denotes firm i's expenditure share on variety v of product k, S^i firm sales, V_k^i is the number of varieties firm i sources of product k and X^i contains additional firm characteristics. According to the theory, we expect that $\gamma_{vk} > 0$ and $\beta_{vk} = \mu_{vk} = 0$.[27] As our dataset contains 9,500 products and 220 countries, this approach would deliver $9,500 * 220 \approx 2,000,000$ predictions.[28] We rather choose a more

[24]This section deals only with foreign varieties. The domestic variety is excluded from this section, as it is the subject of the following one.

[25]Note that these price-adjusted qualities contain prices, qualities and variable transport costs. See equation (4).

[26]It should be noted that this result follows from two assumptions on technology: constant returns to scale and factor neutral productivity. The CES production function provides us with convenient closed form expressions.

[27]Note that the error term u_i in (12) contains the term $\xi_k = \left(V_k\left(\varphi\right)^{-1}\sum_j^{V_k(\varphi)}\xi_{jk}^{\rho-1}\right)^{\frac{1}{\rho-1}}$ (see (11)), which is a function of $V_k\left(\varphi\right)$. Hence, we do not expect (12) to identify the structural parameter, i.e. we do not expect $\gamma_{vk} = 1$. However, since ξ_k is normalized by $V_k\left(\varphi\right)^{-1}$, if the qualities ξ_{vk} were constant within each product, then ξ_k would be constant too and would not depend on $V_k\left(\varphi\right)$. Thus OLS would be consistent in this case.

[28]This number is an upper bound because in practice many product-variety pairs are empty.

Dep. var.	ln expenditure share on top foreign variety [ln] mean = [-0,075] 0,942					
	(1)	(2)	(3)	(4)	(5)	(6)
log Sales	-0.013***		0.003***	0.002***		
	(0.000)		(0.000)	(0.000)		
log No. of Var.		-0.313***	-0.314***	-0.315***	-0.315***	-0.314***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Exporter				-0.001*	0.000	0.000
				(0.000)	(0.000)	(0.000)
Corporate Group				0.000	0.000	0.000
				(0.000)	(0.000)	(0.000)
Foreign Group				0.000	0.000	0.000
				(0.001)	(0.001)	(0.001)
TFP-LP					0.000	
					(0.000)	
TFP-OLS						0.001**
						(0.000)
Product FE	Y	Y	Y	N	N	N
Firm FE	N	N	N	Y	Y	Y
R^2	0.068	0.609	0.610	0.616	0.615	0.615
Observations	2,864,122	2,882,210	2,864,122	2,863,926	2,715,758	2,796,491

Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels. All regressions include year fixed effect and 3 digit industry fixed effects. Columns (1) - (3) contain product fixed effects (8 digit). We weigh observations such that each firm has an equal weight. The dependent variable is $ln(s_k^{max})$, where $s_k^{max} = \max_{v \in V_k} s_k^v$ is the expenditure share on the most popular variety of product k. The number of varieties is the number of countries where product k is sourced from. TFP-OLS is calculated as the residual from an industry-specific OLS regression of log value added on log capital and log employment. TFP-LP is calculated using the method of Levinsohn and Petrin (2003). A firm is member of a foreign group if at least one affiliate or the headquarter is located outside of France. A firm is member of a corporate group if it is controlled by another firm or it has control over at least one affiliate.

Table 5: Determinants of foreign expenditure shares

parsimonious specification, where we focus on the expenditure share of the main variety only. Specifically, we take $s_k^{i,max} = \max_{v \in V_k^i} s_{vk}^i$, that is firm i's expenditure share on its most important variety of product k, as the dependent variable and estimate

$$ln\left(s_k^{i,max}\right) = \alpha_k - \gamma ln\left(V_k^i\right) + \beta ln\left(S^i\right) + \mu X^i + u_k^i,$$

now pooling the data across products.[29] Note that we allow for a product-specific intercept (α_k) but restrict the other coefficients to be the same across products. Table 5 reports the results.

[29]Note that we could also run (12) on the pooled data (across products and varieties). This however does not deliver a very stringent test of the model. Since shares add up to one, the average expenditure on each variety mechanically equals $1/V$, if V varieties of product k are imported. Thus once the number of varieties is controlled for, it might not be surprising that other firm characteristics cease to be important. This is the case empirically. The results are not reported here, but available upon request

The first two columns show that more productive firms, as measured by sales, and firms sourcing more varieties have lower average expenditure shares on their top variety within products. In column (3) we include both sales and the number of varieties sourced and find that while the coefficient on V_k^i hardly changes, the coefficient on sales drops to an economically insignificant level (in absolute value).[30] Hence, the effect of productivity on firms' expenditure shares within products is mainly through its effect via the sourcing strategy as predicted by the theory. In column (4) we add both firm fixed effects and additional firm characteristics as exporting status and membership in a foreign and corporate group. We find that the previous results are confirmed, as the coefficient on the number of varieties sourced hardly changes while the additional firm-level controls are insignificant. Furthermore, the R^2 is basically unchanged, which shows that firm fixed effects, and hence any time-invariant firm characteristics, have little explanatory power.[31] Columns (5) and (6) show that these results are robust to using different measures of productivity.[32]

The main takeaway of the results in Table 5 is that the data does not seem to require firm-level heterogeneity, in particular productivity, to account for the determination of expenditure shares across imported products once the sourcing strategy is controlled for. This result is important, especially for applied theoretical work, because it provides support for the assumption of factor-neutral productivity differences across firms. Note that Kugler and Verhoogen (2010) find support for a complementarity between firm productivity and input quality, which is at odds with factor-neutral productivity. More precisely, they show that more productive firms source products of higher quality. The results reported in Table 5, however, do not concern this extensive margin. We study expenditure shares holding the sourcing strategy fixed. We come back to this discussion below.

Prediction 5. Domestic expenditure shares are decreasing in productivity. However, conditional on the sourcing strategy, productivity is irrelevant. The previous prediction should hold for every variety that a firm sources, in particular the domestic one. In this section we deal separately with the demand for domestic varieties for two reasons. First, as this paper is concerned with firms' import behavior, the domestic variety is of natural importance. Second, we do not observe domestic expenditure shares on the product level but

[30] To see this, note that Table 5 implies that a 10% increase in sales increases the expenditure share on the most popular variety by 0.03%. As the share of the top variety is in the order of 0.9, such change would increase the share to 0.9003.

[31] As a caveat, note that for computational reasons column (4) does not include product fixed effects.

[32] The results are also robust to using shares (as opposed to log shares) as dependent variable. We report the log specification (see (12)) because it follows from the theory.

only for the firm as a whole.[33] Thus the regression of the previous section (see (12)) is not feasible. In our model, the domestic expenditure share at the firm level is given by (see (10))

$$s^D(\varphi) = \sum_{k=1}^{K} s_k(\varphi)\, s_k^D(\varphi) = \sum_{k=1}^{K} \frac{\left(\sum_j^{V_k(\varphi)} \xi_{jk}^{\rho-1}\right)^{\frac{\varepsilon-1}{\rho-1}}}{\sum_{k=1}^{K}\left(\sum_{j=1}^{V_k(\varphi)} \xi_{jk}^{\rho-1}\right)^{\frac{\varepsilon-1}{\rho-1}}} s_k^D(\varphi), \tag{13}$$

where $s_k^D(\varphi)$ is the domestic expenditure share in product k given in (9). Assuming that the domestic expenditure shares do not vary too much across products ($s_k^D(\varphi) \approx s^D(\varphi)$), (13) and (12) suggest to run the regression

$$log\left(s_D^i\right) = \alpha - \gamma ln\left(\overline{V}_i\right) + \beta ln\left(S^i\right) + \mu X^i + u_i, \tag{14}$$

where $\overline{V}_i$ is the average number of varieties firm i imports.[34] Similarly to (11), (13) suggests that domestic expenditure shares should be negatively related to the average number of varieties sourced ($\gamma > 0$). Additionally, more productive firms should have lower domestic shares (since they source more varieties) but conditional on the number of varieties, productivity or any other firm characteristic should be irrelevant ($\beta = \mu = 0$). In other words, the effect of productivity on domestic expenditure shares happens only via the sourcing strategy.

It is important to note that aggregation bias is a potential concern when the regressions are done at the firm level. The reasons is as follows. The main intensive margin prediction of the theory is that firm characteristics should not affect the allocation of expenditure *across* varieties *within* products. At the firm-product level, product fixed effects allow us to isolate this source of variation. This is not possible once the data is aggregated at the firm level. If the quality distribution of varieties differs across products - so that some products feature more concentrated expenditure shares -, and if firms differ in the products they use, a firm level regression may show a spurious correlation between expenditures shares and firm characteristics. To address this concern and for comparability reasons, we re-do the regressions of the previous section at the firm level.[35] Thus, we are able to assess the likely magnitude of

[33] Domestic input expenditure is measured as the difference between total expenditure on wares and inputs and the total value of imports.

[34] To see this more formally, write the domestic expenditure share as

$$ln\left(s_k^D(\varphi)\right) = ln\left(\frac{\xi_{Dk}^{\rho-1}}{\sum_{j\in V_k(\varphi)} \xi_{jk}^{\rho-1}}\right) = -ln\left(\overline{V}(\varphi)\right) + (\rho-1)\left(ln\left(\xi_{Dk}\right) - ln\left(\overline{V}(\varphi)^{-1} \sum_{j\in V_k(\varphi)} \xi_{jk}^{\rho-1}\right)\right),$$

where $\overline{V}(\varphi) = \sum_k s_k(\varphi) V_k(\varphi)$. Under the assumption that $s_k^D(\varphi)$ does not vary across k, the last term in brackets is independent of k and (14) follows.

[35] More precisely, we re-estimate (14) using the log average expenditure share on the most popular variety as the dependent variable.

such aggregation problem.

Table 6 contains the results for the domestic shares (panel A) and for the top foreign shares (panel B). Columns (1) and (2) in Panel A support the first part of prediction 5: both productivity and the average number of varieties have a negative effect on domestic expenditure shares. Column (3), however, shows that productivity matters for domestic expenditure shares above and beyond the sourcing strategy, as the coefficient for sales is positive and significant. Quantitatively, a one standard deviation increase in log sales (holding the sourcing strategy constant) implies an increase in the domestic expenditure share of 7% - which, given that the average domestic shares is about 70%, amounts to 5 additional percentage points. This is in stark contrast with the result for the shares on foreign varieties in Table 5, which shows an economically insignificant coefficient on sales. More importantly, the sales coefficient in the domestic shares regression (see column (3), panel A) is three and half times as high as the one in the foreign shares regression (see column (3), panel B). By comparing panels A and B we are confident that this result is not due to aggregation bias. Note also the large difference in the R^2 between panels. Thus, we find that the sourcing strategy, which according to the model should be a "sufficient statistic" for the pattern of expenditure, has low explanatory power when applied to domestic spending, but explains foreign input demand relatively well. Column (4) adds additional firm level controls and firm fixed effects. While the sales coefficient in panel A drops, it is still significant and is now four times larger than its counterpart in panel B. Moreover, firm level characteristics are economically significant, while this is not the case in panel B. Both being an exporter and being member of a corporate or foreign group reduce the expenditure share on domestic inputs. Finally, the large increase in R^2 from column (3) to (4) in panel A confirms the importance of firm characteristics as determinants of domestic expenditure shares. Columns (5) and (6) report the results using other measures of firm level productivity as robustness measures. While the coefficients on the TFP measures in panel A still exceed the ones in panel B, they are too imprecisely estimated to reach a firm conclusion. However, firm characteristics are still much more important in panel A than in panel B.

Taken together, these findings suggest that the theory is applicable to foreign expenditure shares, but that it is incomplete to account for the pattern of domestic spending. That is, simply relying on factor neutral productivity differences across firms seems insufficient to describe the firm-level evidence on relative domestic spending.[36]

[36]It is interesting to note that Halpern, Koren, and Szeidl (2009), who use a similar demand structure to ours, have also problems to account for the domestic expenditure shares using their structural estimation. At their estimated parameters, the optimal import share is roughly 60% (see their Table 3), which is more than double the share observed in the data (28%).

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A-Domestic	ln expenditure share on domestic inputs					
	[ln] mean = [-0.500] 0.700					
log Sales	-0.019***		0.042***	0.028***		
	(0.001)		(0.001)	(0.006)		
log Avg. No of Var.		-0.295***	-0.359***	-0.162***	-0.156***	-0.156***
		(0.003)	(0.004)	(0.005)	(0.005)	(0.005)
Exporter				-0.042***	-0.037***	-0.036***
				(0.006)	(0.005)	(0.005)
Corporate Group				-0.019***	-0.018***	-0.019***
				(0.006)	(0.006)	(0.006)
Foreign Group				-0.017	-0.016	-0.017
				(0.011)	(0.011)	(0.011)
TFP-LP					0.008*	
					(0.004)	
TFP-OLS						0.005
						(0.006)
Firm FE	N	N	N	Y	Y	Y
R^2	0.041	0.093	0.101	0.763	0.770	0.770
Observations	152,367	154,173	152,367	152,348	143,449	145,143
Panel B-Foreign	ln average expenditure share on top foreign variety					
	[ln] mean = [-0.114] 0.904					
log Sales	-0.032***		0.012***	0.007***		
	(0.000)		(0.000)	(0.001)		
log Avg. No. of Var.		-0.240***	-0.257***	-0.248***	-0.246***	-0.245***
		(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Exporter				-0.002**	-0.001	-0.001
				(0.001)	(0.001)	(0.001)
Corporate Group				0.003**	0.003**	0.003***
				(0.001)	(0.001)	(0.001)
Foreign Group				0.001	0.000	0.000
				(0.002)	(0.002)	(0.002)
TFP-LP					0.003***	
					(0.001)	
TFP-OLS						0.003***
						(0.001)
Firm FE	N	N	N	Y	Y	Y
R^2	0.115	0.614	0.622	0.823	0.823	0.823
Observations	167,733	170,267	167,733	167,711	156,506	158,460

Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels. All regressions include year fixed effects and 3 digit industry fixed effects. The dependent variable in Panel A is the log of the domestic expenditure share. The dependent variable in Panel B is the log of the firm-level average expenditure share on the most popular variety of each product sourced, i.e.$\sum_k s_k^{max}\omega_k$, where ω_k is the expenditure share of product k and $s_k^{max} = \max_{v\in V_k} s_k^v$. "Log Avg. No of Var." is the firm-level average number of varieties across products, $\overline{V} = \sum_k V_k\omega_k$. TFP-OLS is calculated as the residual from an industry-specific OLS regression of log value added on log capital and log employment. TFP-LP is calculated using the method of Levinsohn and Petrin (2003). A firm is member of a foreign group if at least one affiliate or the headquarter is located outside of France. A firm is member of a corporate group if it is controlled by another firm or it has control over at least one affiliate.

Table 6: Domestic and foreign expenditure shares at the firm level

Discussion We consider the results contained in Tables 5 and 6 as the most relevant ones for future research. Whereas Table 5 was in line with foreign spending patterns being uncorrelated with firm characteristics, Table 6 indicates that this is not true concerning the domestic-vs-foreign spending allocation. This means that the relative effective price of foreign inputs varies across firms for a given sourcing strategy. The basic theory we consider, however, rules this out. With CES production structure, the relative price of foreign varieties when sourcing $V = (V_1, ..., V_K)$ varieties is given by

$$P_F(V) = \left(\sum_k^K \left(\sum_{j \neq 1}^{V_k} \xi_{jk}^{\rho-1} \right)^{\frac{\varepsilon-1}{\rho-1}} \right)^{\frac{1}{1-\varepsilon}}. \tag{15}$$

Clearly, $P_F(V)$ depends on productivity as the sourcing strategy V does. But conditional on V, (15) shows that $P_F(V)$ does not vary with φ (or any other firm characteristic) under the assumption that firms face common prices and qualities. Although (15) follows from the CES functional form, it is important to note that this property does not require this assumption. As long as productivity differences across firms are factor-neutral, production has constant returns, and prices and qualities are common, all firms face the same relative price of foreign inputs for a given sourcing strategy. The results in the previous sections show that these assumptions are not appropriate.

While it is interesting that the model is rejected for the domestic-foreign comparison, it is important to learn the direction in which the model should be augmented to be consistent with the data. To do so, we need to interpret the signs of the coefficients of interest in panel A of Table 6. However, once we recognize that productivity has an independent effect on expenditure shares (above and beyond the sourcing strategy), and given that productivity matters for import participation (see Table 2), the results in panel A are likely to be subject to selection bias. For this reason, we re-estimate equation (14) using the standard procedure proposed in Heckman (1979). For greater comparability, we also re-do the regressions in panel B controlling for selection. Table 7 has the results.

Consider first panel A. The results for productivity are qualitatively similar to those reported in Table 6. However, the size of the productivity coefficients is between two and four times as large. This tells us that controlling for selection into importing is important, and that firm productivity is a significant determinant of the demand for domestic inputs.[37] In

[37]These results are consistent with a negative selection bias. To see this consider the following example. Suppose our measure of productivity ($\hat{\varphi}$) was a noisy signal of true productivity, that is $\varphi = \hat{\varphi} + \epsilon$. If firms base their decisions on φ, then high (measured) productivity firms have on average low ϵ realizations, once the sourcing strategy is controlled for.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A-Domestic	ln expenditure share on domestic inputs					
	[ln] mean = [-0.500] 0.700					
log Sales	0.016***		0.070***	0.099***		
	(0.001)		(0.002)	(0.010)		
log Avg. No of Var.		-0.275***	-0.349***	-0.155***	-0.149***	-0.149***
		(0.004)	(0.004)	(0.006)	(0.006)	(0.006)
Exporter				0.143***	0.128***	0.129***
				(0.021)	(0.020)	(0.020)
Corporate Group				0.014*	0.013*	0.010
				(0.008)	(0.008)	(0.008)
Foreign Group				0.028**	0.024*	0.023*
				(0.013)	(0.013)	(0.013)
TFP-LP					0.028***	
					(0.006)	
TFP-OLS						0.028***
						(0.007)
Inv. Mill's Ratio	0.176***	0.068***	0.147***	5.203***	4.781***	4.770***
	(0.004)	(0.003)	(0.004)	(0.557)	(0.551)	(0.560)
Firm FE	N	N	N	Y	Y	Y
Observations	846,326	846,326	846,326	846,326	627,392	628,414
Panel B-Foreign	ln average expenditure share on top foreign variety					
	[ln] mean = [-0.114] 0.904					
log Sales	-0.027***		0.013***	0.007***		
	(0.000)		(0.000)	(0.002)		
log Avg. No of Var.		-0.243***	-0.258***	-0.245***	-0.244***	-0.243***
		(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Exporter				(0.001)	(0.002)	(0.002)
				-0.004	-0.004	-0.004
Corporate Group				0.003*	0.003*	0.003*
				(0.002)	(0.002)	(0.002)
Foreign Group				0.003	0.002	0.002
				(0.003)	(0.003)	(0.003)
TFP-LP					0.003***	
					(0.001)	
TFP-OLS						0.003**
						(0.001)
Inv. Mill's Ratio	0.027***	-0.009***	0.006***	0.035	-0.030	-0.033
	(0.001)	(0.001)	(0.001)	(0.103)	(0.107)	(0.108)
Firm FE	N	N	N	Y	Y	Y
Observations	858,394	858,394	858,394	858,394	637,780	639,016

Notes: This table reports the second stage of the estimator proposed by Heckman (1979). Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels. All regressions include year fixed effects and 3 digit industry fixed effects. The dependent variable is the log of the domestic expenditure share. "Log Avg. No of Var." is the firm-level average number of varieties across products, $\overline{V} = \sum_k V_k \omega_k$. TFP-OLS is calculated as the residual from an industry-specific OLS regression of log value added on log capital and log employment. TFP-LP is calculated using the method of Levinsohn and Petrin (2003). A firm is member of a foreign group if at least one affiliate or the headquarter is located outside of France. A firm is member of a corporate group if it is controlled by another firm or it has control over at least one affiliate. The excluded instrument in the first stage is the lagged import status (see Table 2).

Table 7: Domestic expenditure shares controlling for selection into importing

particular, column (1) shows that more productive firms spend relatively more on domestic inputs, despite them sourcing more varieties (as shown in prediction 2). The coefficient increases fivefold once we control for the sourcing strategy in column (3) indicating that the coefficient in column (1) captures two effects. Productivity decreases the expenditure share on domestic inputs via its positive effect on the number of varieties, but has also a positive direct effect. This result is confirmed with the other two measures of firm level productivity which are positive, sizable and highly significant. Panel B of Table 7 confirms that productivity does not have a direct effect on expenditure shares of foreign varieties. After controlling for selection, the coefficients are virtually identical to the ones in panel B of Table 6, and the inverse Mill's ratio is small and ceases to be significant once firm fixed effects are included.

Taken together, these results show that, for a given sourcing strategy, productivity does not affect relative spending across imported products but increases the share of domestic inputs relative to foreign ones. The latter is in line with the hypothesis that firm productivity and input quality are complements (see Kugler and Verhoogen (2010)), if French inputs are of relatively higher quality than foreign ones (Hummels and Klenow, 2005). However note that this type of complementarity can only be part of the story. As foreign inputs differ in their quality[38], such complementarity would imply that more productive firms spend relatively more on varieties of high quality. This is inconsistent with the results in Table 5.

Let us now briefly discuss two directions how the basic framework could be amended to accommodate these findings. A first route would abandon the reliance on a single source of heterogeneity (productivity) and introduce a second characteristic, varying at the firm level. To be consistent with Tables 5 *and 6,* this additional firm characteristic would need to affect the relative price of foreign varieties viz-a-viz domestic inputs but should keep the relative efficiencies of foreign varieties unaffected. Suppose for example that the price-adjusted quality of foreign variety j of product k for firm i was given by $\tilde{\xi}^i_{jk} = \frac{\eta_{jk}}{p_{jk}(1+\tau_{jk})}\mu^i = \xi_{jk}\mu^i$ where μ^i denotes the additional firm characteristic. Then (15) shows that firm $i's$ relative price of foreign inputs is given by

$$\tilde{P}_F\left(V,\mu^i\right) = \left(\sum_k^K\left(\sum_{j\neq 1}^{V_k}\xi_{jk}^{\rho-1}\right)^{\frac{\varepsilon-1}{\rho-1}}\right)^{\frac{1}{1-\varepsilon}}(\mu^i)^{-1} = \frac{P_F\left(V\right)}{\mu^i}.$$

Due to the multiplicative formulation, the relative efficiency of foreign varieties viz-a-viz each other will be unaffected, but "efficient importers" (i.e. high-μ firms) will have a comparative

[38]When viewed through the lens of the model, the concentration of firms' expenditure shares across products and varieties documented above (Facts 3 and 4) are indicative of substantial price-adjusted quality differences (see (9)).

advantage in sourcing internationally. This would generate a positive correlation between productivity and the domestic expenditure share conditional on the sourcing strategy V.[39] Note that this specification is a variation of the mechanism in Kugler and Verhoogen (2010).

A second route would follow the methodological approach of Arkolakis (2008). In his work, exporting firms have to use resources to attract consumers. As more productive firms chose to cater to a bigger market, the reduced form of his model looks *as if* more productive firms face higher fixed costs of exporting. The analogy in our model would be a mechanism by which firms' behavior is reflected in the final price-adjusted qualities in such a way that more productive firms end up facing a higher relative price of foreign inputs. More theoretical and empirical research is needed to inform us about the underlying reason for the "excess" firm-level heterogeneity in domestic-foreign spending patterns.

5 Conclusion

The purpose of this paper is to propose a simple model of import demand, to test its qualitative predictions using firm-level data, and thus to isolate the aspects of the theory which need further refinements. The model is standard in that it relies on two core assumptions: productivity is the single source of firm-level heterogeneity and prices, fixed costs and input qualities are common across firms. The model predicts that more productive firms select into importing, import more products, and import more varieties of each product. Using a comprehensive dataset of French manufacturing firms, we show that all of these extensive margin predictions are borne out. On the intensive margin, the key prediction is that more productive firms should have lower expenditure shares on each variety of each product, but that, conditional on the firm's sourcing strategy, productivity (and in general any firm characteristic) should be irrelevant. While we do find evidence supporting this prediction among expenditure shares across foreign inputs, we reject the prediction for the domestic vs. foreign expenditure pattern. After controlling for the sourcing strategy, we find that more productive firms spend a higher share of their budget on domestic inputs. These results not only show that the core assumptions of the standard model need to be amended, but also suggest the particular direction in which to do so. Further research is required to achieve a better understanding of import demand.

[39]Such a model would generate the empirical correlation between expenditure shares and firm-level productivity regardless of the correlation between φ and μ on the firm level. Efficient importers (high-μ firms) will have a lower productivity cut-off to start importing and will import more products and more varieties conditional on productivity. Hence, conditional on the sourcing strategies, high-μ firms will be (on average) less productive than their low-μ counterparts and will have lower domestic expenditure shares. Hence, conditional on the sourcing strategy, productivity and domestic expenditure shares are positively correlated irrespective of the cross-sectional distribution of the random vector (φ, μ).

6 Appendix

6.1 Data Description

Our main data set stems from the information system of the French custom administration (DGDDI) and contains the universe of import and export flows by French manufacturing firms. The data is collected at the 8-digit (NC8) level and a firm located within the French metropolitan territory must report this detailed information as long as the following criteria are met. Within EU imports, have to be reported as long as the firm's annual trade value exceeds 150,000 Euros. If that threshold is not met, firms can choose to report under a simplified scheme. However, in practice, many firms under that threshold report the detailed information. For imports from outside the EU, all shipments must be reported to the custom administration as this data is used to calculate the value added tax in all cases. The conditions are more stringent for exports. For within EU exports, all shipments must be reported to the custom administration. For exports outside the EU, reporting is required if the exported value to exceeds 1,000 Euros or weighs more than a ton.

The attractive feature of the French data is the presence of unique firm identifiers (the SIREN code), which is available in all French administrative files. Hence, various other datasets can be matched to the trade data at the firm level. To learn about the characteristics of the firms in our sample we employ fiscal files.[40] Sales are deflated using price indices of value added at the 3 digit level obtained from the French national accounts. To measure the expenditure on domestic inputs, we substract the total import value from the total expenditure on wares and inputs reported in the fiscal files. Capital, used for the TFP estimation, was computed using a permanent inventory method. The series were initialized with the deflated value of assets reported in the first year of reported fiscal account (1995). We then used the reported investment expenditure, which we deflated with an investment price index available from the French national accounts. We assumed a depreciation rate of 10%.

Additionally we use the French business registers (SIRENE files), created by the Firm Demography Department of the French National Institute of Statistics (INSEE). The SIRENE files report the yearly creation and destruction of French firms and provide us with information about firm age and legal status. Finally we incorporate information on the ownership structure from the LIFI/DIANE (BvDEP) files. These files are constructed at INSEE using a yearly survey (LIFI) describing the structure of ownership of all of the French firms in the private

[40]The firm level accounting information is retrieved from two different files: the BRN ("Bénéfices Réels Normaux") and the RSI ("Régime Simplifié d'Imposition"). The BRN contains the balance sheet of all firms in the traded sectors with sales above 730,000 Euros. The RSI is the counterpart of the BRN for firms with sales below 730,000 Euros. Although the details of the reporting differs, for our purpose these two data sets contain essentially the same information. Their union covers nearly the entire universe of all French firms.

	Full Sample	Importers	Non-Importers	Exporters	Non-Exporters
Employment	18.18	89.72	5.71	78.89	5.9
Sales	4342.35	23646.8	969.15	20842	1035.52
Sales / Worker	104.42	157.62	92.3	150.46	92.3
Capital / Worker	33.74	44.28	31.29	42.45	31.38
Inv. / Worker	2.9	4.13	2.6	3.91	2.61
Inputs (mat.)	0.2	0.3	0.18	0.28	0.18
Import share	0.04	0.3	0	0.2	0.01
Share of Importers	0.15	1	0	0.66	0.05
Share of Exporters	0.17	0.75	0.07	1	0
Firm Age	14.38	19.45	13.51	19.54	13.36
Foreign owned	0.02	0.13	0.01	0.11	0.01
Foreign Group	0.05	0.23	0.01	0.2	0.02
Labor Productivity	43.783	56.496	40.726	56.023	40.405
TFP-LP	28.135	30.255	27.788	29.23	27.999
TFP-OLS	25.412	26.519	25.251	26.61	25.189
Number of Firms	259,602	31,022	228,580	34,527	225,075

Notes: Sales, wages, expenditures on imports or exports are all expressed in 2000 prices using a 3-digit industry level price deflator. Our capital measure is the book value reported in firms' balance sheets ("historical cost"). We measure employees by occupation. Skilled workers are engineers, technicians and managers, workers of intermediate skills are skilled blue and white collars and low skilled workers are members of unskilled occupations. A firm is foreign owned, if the controlling entity is a foreign company. A firm is member of a foreign group if at least one affiliate or the headquarter is located outside of France.

Table 8: Characteristics of importers, exporters and domestic firms.

sector whose financial investments in other firms (participation) are higher than 1.2 million Euros or having sales above 60 million Euros or more than 500 employees. This survey is complemented with the information about ownership structure available in the DIANE (BvDEP) files, which are constructed using the annual mandatory reports to commercial courts, and with the register of firms that are controlled by the State.

Using these datasets, we construct a non-balanced panel dataset spanning the period from 2001 to 2006. Some basic characteristics of importing and non-importing firms are contained in Table 8. For comparison, we also report the results for exporting firms.

6.2 A necessary condition for hierarchical sourcing

We are going to describe the condition for $V(\varphi)$ to be increasing in φ, i.e. we are going to characterize the conditions such that a variety is sourced by firm φ' whenever it is imported by firm $\varphi < \varphi'$. Consider variety j of product k and suppose this is sourced by firm φ but not by φ'. Let $V_{-jk}(\varphi)$ be the vector of varieties sourced by φ excluding jk and $V_{+jk}(\varphi')$ the vector of varieties of φ' when jk is added. Then

$$A\varphi^{\sigma-1}\left[c\left(V\left(\varphi\right)\right)^{1-\sigma}-c\left(V_{-jk}\left(\varphi\right)\right)^{1-\sigma}\right] \quad > \quad f_{jk} > A\varphi'^{\sigma-1}\left[c\left(V_{+jk}\left(\varphi'\right)\right)^{1-\sigma}-c\left(V\left(\varphi'\right)\right)^{1-\sigma}\right].$$

Hence,

$$\left(\frac{\varphi'}{\varphi}\right)^{\sigma-1} < \frac{c\left(V\left(\varphi\right)\right)^{1-\sigma}-c\left(V_{-jk}\left(\varphi\right)\right)^{1-\sigma}}{c\left(V_{+jk}\left(\varphi'\right)\right)^{1-\sigma}-c\left(V\left(\varphi'\right)\right)^{1-\sigma}} = \left(\frac{c\left(V\left(\varphi\right)\right)}{c\left(V\left(\varphi'\right)\right)}\right)^{1-\sigma}\frac{\left(\frac{c(V(\varphi))}{c(V_{-jk}(\varphi))}\right)^{1-\sigma}-1}{\left(\frac{c(V_{+jk}(\varphi'))}{c(V(\varphi'))}\right)^{1-\sigma}-1}$$

A first order approximation around $\frac{c(V(\varphi))}{c(V_{-jk}(\varphi))}=1$ yields

$$\left(\frac{c\left(V\left(\varphi\right)\right)}{c\left(V_{-jk}\left(\varphi\right)\right)}\right)^{1-\sigma}-1 \approx \frac{1}{1-\sigma}\frac{c\left(V\left(\varphi\right)\right)-c\left(V_{-jk}\left(\varphi\right)\right)}{c\left(V_{-jk}\left(\varphi\right)\right)} = \frac{1}{1-\sigma}\Delta_{jk}\left(\varphi\right),$$

where $\Delta_{jk}\left(\varphi\right)$ is the percentage efficiency gain for firm φ to import variety jk. We then require that

$$\frac{\Delta_{jk}\left(\varphi\right)}{\Delta_{jk}\left(\varphi'\right)} > \left(\frac{\varphi'}{\varphi}\right)^{\sigma-1}\left(\frac{c(V\left(\varphi\right))}{c\left(V\left(\varphi'\right)\right)}\right)^{\sigma-1} = \left(\frac{\varphi'/c\left(V\left(\varphi'\right)\right)}{\varphi/c\left(V\left(\varphi\right)\right)}\right)^{\sigma-1} = \frac{S\left(\varphi'\right)}{S\left(\varphi\right)} > 1, \qquad (16)$$

where $S\left(\varphi\right)$ denotes firm φ's sales. Hence, firm φ' also sources kj whenever the relative efficiency gain for firm φ does not exceed the relative sales, i.e. whenever the efficiency gains are not too unequal.

References

ACEMOGLU, D. (1998): "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality," *Quarterly Journal of Economics*, 113(4), 1055–1089.

——— (2003): "Patterns of Skill Premia," *Review of Economic Studies*, 70, 199–230.

——— (2007): "Equilibrium Bias of Technology," *Econometrica*, 175, 1371–1410.

ACEMOGLU, D., AND U. AKCIGIT (forthcoming): "Intellectual Property Rights Policy, Competition and Innovation," *Journal of the European Economic Association.*

ACEMOGLU, D., P. ANTRAS, AND E. HELPMAN (2007): "Contracts and Technology Adoption," *American Economic Review*, 97(3), 916–943.

ACEMOGLU, D., S. JOHNSON, AND J. ROBINSON (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91(5), 1369–1401.

ACEMOGLU, D., AND J. VENTURA (2002): "The World Income Distribution," *Quarterly Journal of Economics*, 117(2), 659–694.

ACEMOGLU, D., AND F. ZILIBOTTI (1999): "Information Accumulation in Development," *Journal of Economic Growth*, 4(1), 5–38.

AGHION, P., AND P. HOWITT (1992): "A Model of Growth through Creative Destruction," *Econometrica*, 60(2), 323–351.

AMITI, M., AND J. KONINGS (2007): "Trade Liberalization, Intermediate Inputs and Productivity," *American Economic Review*, 97(5), 1611–1638.

ARKOLAKIS, C. (2008): "Market Penetration Costs and the New Consumer Margin in International Trade," NBER Working paper.

ARMENTER, R., AND M. KOREN (2008): "A Balls-and-Bins Model of Trade," Working Paper.

ARMINGTON, P. S. (1969): "A Theory of Demand for Products Distinguished by Place and Production," *IMF Staff Paper*, 16, 159–178.

ATKESON, A., AND A. BURSTEIN (2008): "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 98(5), 1998–2031.

——— (2010): "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, 118(3), 433–484.

BANERJEE, A., AND E. DUFLO (2005): "Growth Theory through the Lens of Development Economics," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf, vol. 1A, chap. 7, pp. 473–552. Elsevier B.V.

BANERJEE, A., AND B. MOLL (2010): "Why Does Misallocation Persist?," *American Economic Journal: Macroeconomics*, 2(1), p. 189–206.

BANERJEE, A., AND A. NEWMAN (1998): "Information, the Dual Economy, and Development," *Review of Economic Studies*, 65, 631–653.

BARSEGHYAN, L., AND R. DICECIO (2009): "Entry Costs, Misallocation and Cross-Country Income and TFP Differences," Working paper, Federal Reserve Bank of St. Louis.

BARTELSMAN, E. J., J. C. HALTIWANGER, AND S. SCARPETTA (2009): "Cross-Country Differences in Productivity: The Role of Allocation and Selection," Working paper 15490, NBER.

BARTON, B. (1948): *The problem of 12 million German refugees in today's Germany.* American Friends Service Commitee, Philadelphia.

BASU, S., AND J. G. FERNALD (2002): "Aggregate Productivity and Aggregate Technology," *European Economic Review*, 46, 963–991.

BERNARD, A., J. EATON, J. B. JENSEN, AND S. KORTUM (2003): "Plants and Productivity in International Trade," *American Economic Review*, 93(4), 1268–1290.

BERNARD, A. B., J. JENSEN, AND P. SCHOTT (2009): "Importers, exporters and multinationals: A portrait of firms in the U.S. that trade goods," in *Producer Dynamics: New Evidence from Micro Data*, ed. by T. Dunne, J. Jensen, and M. Roberts. University of Chicago Press.

BERNARD, A. B., AND J. B. JENSEN (1999): "Exceptional Exporter Performance: Cause, Effect, or Both?," *Journal of International Economics*, 47(1), 1–25.

BERNARD, A. B., J. B. JENSEN, S. REDDING, AND P. SCHOTT (2007): "Firms in International Trade," *Journal of Economic Perspectives*, 3(21), 105–130.

BLALOCK, G., P. J. GERTLER, AND D. I. LEVINE (2008): "Financial Constraints on Investment in an Emerging Market Crisis," *Journal of Monetary Economics*, 55(3), 568–591.

BORJAS, G. J. (1999): "The Economic Analysis of Immigration," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3, chap. 28, pp. 1697–1760. Elsevier.

——— (2003): "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *Quarterly Journal of Economics*, 118(4), 1335–1374.

BRAKMAN, S., H. GARRETSON, AND M. SCHRAMM (2004): "The Strategic Bombing of German Cities during WWII and its Impact in City Growth," *Journal of Economic Geography*, 4(2), 201–18.

BRODA, C., J. GREENFIELD, AND D. E. WEINSTEIN (2006): "From Groundnuts to Globalization: A Structural Estimate of Trade and Growth," NBER Working paper.

BUERA, F., J. KABOSKI, AND Y. SHIN (2011): "Finance and Development: A Tale of Two Sectors," *American Economic Review*, 101(5), 1964–2002.

CARD, D. (1990): "The Impact of the Marial Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review.*

CASELLI, F. (2005): "Accounting for Cross-Country Income Differences," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf, vol. 1A, chap. 9, pp. 679–741. Elsevier B.V.

CASELLI, F., AND W. J. COLEMAN II (2001): "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy*, 109(3), 584–616.

CHANEY, T. (2008): "Distorted Gravity: the Intensive and Extensive Margins of International Trade.," *The American Economic Review*, 98(4), 1707–1721.

CHARI, V., AND H. HOPENHAYN (1991): "Vintage Human Capital, Growth and the Diffusion of New Technology," *Journal of Political Economy*, 99, 1142–65.

CHARI, V., P. J. KEHOE, AND E. R. MCGRATTAN (2007): "Business Cycle Accounting," *Econometrica*, 75(3), 781–836.

CLEMENTI, G. L., AND H. HOPENHAYN (2006): "A Theory of Financing Constraints and Firm Dynamics," *The Quarterly Journal of Economics*, 121(1), 229–265.

COLLARD-WEXLER, A., J. ASKER, AND J. DE LOECKER (2011): "Productivity Volatility and the Misallocation of Resources in Developing Economies," Discussion paper, NBER Working Paper 17175.

DAVIS, D. R., AND D. E. WEINSTEIN (2002): "Bones, Bombs and Break Points: The Geography of Eonomic Activity," *American Economic Review*, 92(5), 1269–89.

DE LOECKER, J. (2011): "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity," *Econometrica*, 79(5), 1407–1451.

DE LOECKER, J., AND F. WARZYNSKI (forthcoming): "Markups and Firm-Level Export Status," *American Economic Review.*

DE ZAYAS, A.-M. (1993): *The German Expellees: Victims in War and Peace.* Palgrave Macmillan.

DINARDO, J., AND J.-S. PISCHKE (1997): "The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too?," *Quarterly Journal of Economics*, 112(1), 291–303.

DJANKOV, S., R. LA PORTA, F. LOPEZ-DE-SILVAES, AND A. SHLEIFER (2002): "The Regulation of Entry," *Quarterly Journal of Economics*, 117(1), 1–37.

DUSTMANN, C., J. LUDSTECK, AND U. SCHOENBERG (2009): "Revisiting the German Wage Structure," *Quarterly Journal of Economics*, 124(2), 843–881.

EATON, J., S. KORTUM, AND F. KRAMARZ (2004): "Dissecting Trade: Firms, Industries, and Export Destinations," *The American Economic Review*, Vol. 94(2), 150–154.

——— (2008): "An Anatomy of International Trade: Evidence from French Firms," Working Paper.

EDMOND, C., V. MIDRIGAN, AND D. Y. XU (2011): "Competition, Markups and the Gains from International Trade," Working paper, Federal Reserve Bank of Minneapolis.

EPIFANI, P., AND G. GANCIA (2011): "Trade, Markup Heterogeneity and Misallocations," *Journal of International Economics*, 83(1), 1–13.

FATTAL JAEF, R. N. (2011): "Entry, Exit and Misallocation Frictions," Discussion paper.

FEENSTRA, R. C. (1998): "Integration of Trade and Disintegration of Production in the Global Economy,," *Journal of Economic Perspectives*, pp. 31–50.

FOSTER, J., J. C. HALTIWANGER, AND C. SYVERSON (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, 98(1), 394–425.

FRIEDBERG, R. (1997): "The Impact of Mass Migration on the Israeli Labor Market," *Quarterly Journal of Economics*, 116(4).

GANCIA, G., AND F. ZILIBOTTI (2009): "Technological Change and the Wealth of Nations," *Annual Review of Economics*, 1, 93–120.

GOLDBERG, P., A. KHANDELWAL, N. PAVCNIK, AND P. TOPALOVA (2009): "Trade Liberalization and New Imported Inputs," *American Economic Review (P&P)*, 99(2), 494–500.

GROSSER, T., AND S. SCHRAUT (2001): *Fluechtlinge und Heimatvertriebene in Wuerttemberg-Baden nach dem Zweiten Weltkrieg: Dokumente und Materialien zu ihrer Aufnahme und Eingliederung.* Mannheim.

GROSSMAN, G., AND E. HELPMAN (1991): "Quality Ladders in the Theory of Growth," *Review of Economic Studies*, 58(1), 43–61.

HALL, R., AND C. JONES (1999): "Why Do Some Countries Produce So Much More Ouput Per Worker Than Others?," *Quarterly Journal of Economics*, 114(1), 83–116.

HALPERN, L., M. KOREN, AND A. SZEIDL (2009): "Imported Inputs and Productivity," Working Paper.

HANLON, W. W. (2011): "Necessity if the Mother of Invention: Input Supplies and Directed Technical Change," Working Paper.

HECKMAN, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–61.

HOPENHAYN, H. (1992): "Entry, Exit and Firm Dynamics in Long-Run Equilibrium," *Econometrica*, 60(5), 1127–1150.

HSIEH, C.-T., AND P. KLENOW (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124(2), 771–807.

——— (2011): "The Life-Cycle of Plants in India and Mexico," Discussion paper, Working Paper.

HUMMELS, D., J. ISHII, AND K.-M. YI (2001): "The nature and growth of vertical specialization in world trade," *Journal of International Economics*, 54(1), 75–96.

HUMMELS, D., AND P. KLENOW (2005): "The Variety and Quality of a Nation's Exports," *The American Economic Review*, 95(3), 704–723.

ISAAC, J. (1948): *German Refugees in the U.S. Zone 1948/49.*

JEONG, H., AND R. TOWNSEND (2007): "Sources of TFP Growth," *Journal of Economic Theory*, 32(1), 179–221.

KLENOW, P., AND A. RODRIGUEZ-CLARE (1997): "The Neoclassical Revival in Growth Economics: Has it Gone Too Far?," in *NBER Macroeconomics Annual*, vol. 12, pp. 73–103.

KLETTE, T. J., AND S. KORTUM (2004): "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 112(5), 986–1018.

KORNRUMPF, M. (1979): *In Bayern angekommen. Die Eingliederung der Vertriebenen: Namen, Daten, Fakten.* Olzog, Muenchen.

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

——— (1991): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99(3), 483–499.

KUGLER, M., AND E. VERHOOGEN (2010): "The Quality-Complementarity Hypothesis: Theory and Evidence from Colombia," NBER Working paper.

——— (forthcoming): "Plants and Imported Inputs: New Facts and an Interpretation," *American Economic Review (P&P)*.

LA PORTA, R., AND A. SHLEIFER (2009): "The Unofficial Economy and Economic Development," in *Brookings Papers on Economic Activity*, pp. 275–352.

LAGAKOS, D., AND M. E. WAUGH (2011): "Selection, Agriculture and Cross-Country Productivity Differences," Working Paper.

LAGOS, R. (2006): "A Model of TFP," *Review of Economic Studies*, 73(4), 983–1007.

LENTZ, R., AND D. T. MORTENSEN (2008): "An Empirical Model of Growth through Product Innovation," *Econometrica*, 76(6), 1317–1373.

LEVINSOHN, J., AND A. PETRIN (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–341.

LEWIS, A. (1954): "Economic Development with Unlimited Supplies of Labor," *The Manchester School*, 22(2), 131–191.

LEWIS, E. (2011): "Immigration, Skill Mix and Capital-Skill Complementarity," *Quarterly Journal of Economics*, 126(2), 1029–1069.

LUCAS, R. E. (1978): "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 9(2), 508–523.

——— (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22(1), 3–42.

——— (1990): "Why Doesn't Capital Flow From Rich to Poor Countries," *American Economic Review*, 80(2), 92–96.

LUETTINGER, P. (1989): *Integration der Vertriebenen: eine empirische Analyse.* Campus, Frankfurt/Main; New York.

LUTTMER, E. G. J. (2010): "Models of Growth and Firm Heterogeneity," *Annual Review of Economics*, 2, 547–576.

MATSUYAMA, K. (1992): "Agricultural Productivity, Comparative Advantage and Economic Growth," *Journal of Economic Theory*, 58(2), 317–334.

——— (1995): "Complementarities and Cumulative Processes in Models of Monopolistic Competition," *Journal of Economic Literature*, 33(2), 701–729.

MAYER, T., M. MELITZ, AND G. OTTAVIANO (2010): "Market size, Competition, and the Product Mix of Exporters," mimeo.

MCKINSEY-GLOBAL-INSTITUTE (2001): "India: The Growth Imperative," .

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

MELITZ, M., E. HELPMAN, AND S. YEAPLE (2004): "Export Versus FDI with Heterogeneous Firms," *The American Economic Review*, 94(1), 300–316.

MELITZ, M., AND G. OTTAVIANO (2008): "Market Size, Trade and Productivity," *Review of Economic Studies*, 75(1), 295–316.

MIDRIGAN, V., AND D. Y. XU (2010): "Finance and Misallocation: Evidence From Plant-Level Data," NBER Working Paper 15647.

MOLL, B. (2010): "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?," Working Paper.

MURPHY, K. M., A. SHLEIFER, AND R. W. VISHNY (1989): "Industrialization and the Big Push," *Quarterly Journal of Economics*, 106, 503–530.

NELLNER, W. (1959): "Grundlagen und Hauptergebnisse der Statistik," in *Die Vertriebenen in Westdeutschland: ihre Eingliemng und ihr Einluss auf die Gesellschaft, Wirtschaft, Politik und Geistesleben*, ed. by E. Lember, and F. Edding. Verlag Ferdinand Hirt.

OFFICE OF THE MILITARY GOVERNMENT FOR GERMANY (1945): "A year of Potsdam: The German economy since surrender," Discussion paper, Military Governor.

——— (1947): "Population of the US Zone of Germany (Summary)," Discussion paper, Civil Administration Divison.

——— (1949): "Information Bulletin No. 169," Discussion paper, Provisional Office of Administration, Reports & Statistics Branch, US Army.

PARENTE, S. L., AND E. C. PRESCOTT (1994): "Barriers to Technology Adoption and Development," *Journal of Political Economy*, 102(2), 298–321.

PAVCNIK, N. (2002): "Trade Liberalization, Exit and Productivity Improvements: Evidence from Chilean Plants," *Review of Economic Studies*, 69(1), 245–276.

PETERSEN, W. (1964): "Review," *American Sociological Review*, 29(3), 419–421.

REICHLING, G. (1958): *Die Heimatvertriebenen im Spiegel der Statistik.* Duncker & Humboldt, Berlin.

RESTUCCIA, D., AND R. ROGERSON (2008): "Policy Distortions and Aggregate Productivity with Heterogenous Establishments," *Review of Economic Dynamics*, 11(4), 707–720.

ROY, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–146.

SPITZ-OENER, A. (2006): "Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure," *Journal of Labor Economics*, 24(3), 235–270.

STATISTISCHES BUNDESAMT (1953): *Statistisches Taschenbuch über die Heimatvertriebenen in der Bundesrepublik Deutschland und in West-Berlin.* Statistisches Bundesamt, Wiesbaden.

——— (1955): "Die Vertriebenen und Fluechtlinge in der Bundesreublik Deutschland 1946-1953," in *Statistik der Bundesrepublik Deutschland*, vol. 114, Stuttgart. Kohlhammer.

SUTTON, J. (1997): "Gibrat's Legacy," *Journal of Economic Literature*, 35, 40–59.