THÈSE de DOCTORAT

de l'UNIVERSITÉ TOULOUSE CAPITOLE

Présentée et soutenue par

Monsieur Oscar FENTANES TELLEZ

Le 14 juin 2024

Essais sur la dynamique des entreprises et le développement

École doctorale : **TSE Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche : **TSE-R - Toulouse School of Economics - Recherche**

Thèse dirigée par *Monsieur Fabrice COLLARD et Monsieur Christian HELLWIG*

Composition du jury

Rapporteur : M. Neira JULIAN
Rapporteur : M. Dupaigne MARTIAL
Examinateur : M. Charles Brendon
Directeur de thèse : M. Fabrice COLLARD

Essays on Firm Dynamics and Development

Ph.D. Thesis

Oscar Eduardo Fentanes Tellez[1]

Toulouse School of Economics

June 2024

[1] Address: 1, Esplanade de l'Université, 31080 Toulouse, France; email: o.fentanes.t@gmail.com

Acknowledgements

I would like to first express my gratitude to Fabrice Collard and Christian Hellwig for their amazing guidance and support as my professors and advisors. I am deeply indebted to Patrick Fève, Eugenia Gonzalez-Aguado, and Stéphane Straub for their invaluable suggestions at every stage of my research projects. Highly significant were the contributions of other past and current professors from the Macro group, including, Tiziana Assenza, Nicolas Werquin, Sumudu Kankanamge, Martial Dupaigne, Andreas Schaab and Charles Brendon.

I cannot thank my coauthors enough. First, Jonas Gathen, for being the best friend and coauthor. His energy for both research and hikes will always inspire me. I am especially indebted to Matias Busso and Santiago Levy, my mentors for more than a decade. They motivated me to start the doctoral program and have closely supported me since the beginning. I must mention Natalia Volkow, whose professionalism and integrity as the head of INEGI's Microdata Laboratory made my research possible.

Many thanks to my fellow Ph.D. students, especially to my dear friends Sébastien Montpetit and Ignacio Hernández, whose conversations at the university restaurant and Fantastic France (FF) anecdotes made the Ph.D. journey so much fun. To Chanwoo Kim, for our many thought-provoking existential discussions and his endless supply of snacks. And to Basile, Oscar, Anaïs, Léa, Fede, Luisa, Johanna, Mudit, Gyung Mo, and Nicolas – I am greatful for their kindness and intellect. They were the best cohort I could have asked for.

No less important for my success were my friends in Toulouse: Thi, Lukas, Camille, and Moises, with whom I shared many dinners and conversations; and my friends in Mexico, Juan, Isaac, and Pablo, who gave meaning to my summer work trips to Mexico City. Finally, I thank from the bottom of my heart my mother, Rosa, to whom I owe all I have accomplished; and my partner, Florence, for her deep love and for showing me the beauty of Ariège.

Summary

Firms and their entry, exit, and productivity growth processes are at the heart of economic development. This thesis presents three essays on how key elements of the economic environment shape firm dynamics in developing countries, including infrastructure, adjustment frictions, and regulations.

In the first chapter, Matias Busso and I study what determines the aggregate and regional effects of new transportation infrastructure. A key overlooked channel is the role that infrastructure policy plays in changing the incentives of firms to enter, exit, and grow – in turn generating endogenous changes in local productivity. The chapter documents and quantifies the importance of this channel by using detailed Mexican microdata and a spatial general-equilibrium model incorporating firm dynamics. Leveraging random delays in the construction of highways, we empirically show that productivity grows in places with better transportation infrastructure. Firms play a critical role in driving these results: highways increase firms' size, entry rates, survival rates, and total factor productivity. Then, by calibrating our model on census data between 1998 and 2018, we find that new highways over this period increased welfare and income by half a percent, similar to its costs in terms of GDP. Moreover, we find substantial spatial reallocation of workers and production. Nearly half of these effects are explained by endogenous changes in local productivity, which is driven by firm dynamics.

In the second chapter, co-authored with Jonas Gathen, we focus on the drivers of growth miracles. We argue that growth miracles are driven by a fundamental race: as the economy tries to catch up to its steady state, changes in the economic environment move the steady state itself and provide new potential for catch-up growth. We quantify this race over the course of development using 40 years of plant-level manufacturing panel data from Indonesia and a structural model of plant dynamics. We estimate the model on the micro data along the observed growth path without assuming that the economy is ever at a steady state. While catch-up growth starting from initial conditions in 1975 accounts for 42% of Indonesia's subsequent industrialization, new changes in the economy induce new catch-up growth. In the end, the economy is in a never-ending race where it never catches up to its full potential.

In the third chapter, Santiago Levy and I study a common growth paradox in developing countries, where fast industrialization might be coupled with low aggregate productivity growth. We argue that the paradox can be explained by two opposing forces. On the one hand, governments introduce policies that promote growth, such as trade liberalization, competition agencies, and regulatory bodies; and on the other hand, policies segment the economy into formal and informal sectors. To shed light on this outcome, we construct a 20-year establishment-level panel dataset for Mexico, a country where manufacturing exports grew from seven to 33% of GDP, but labor informality barely changed, firm informality increased, and TFP growth was negative. We find that many high-productivity formal firms exit; surviving firms' size and productivity hardly grow, and many informalize; entrants are less productive than survivors, mostly because of large informal entry. Finally, we show that while manufacturing performs better, its contribution to TFP is modest because informality persists in the sector; and despite spectacular export growth, the country is now de-industrializing.

Résumé

Les entreprises et leurs processus d'entrée, de sortie et de croissance de la productivité sont au cœur du développement économique. Cette thèse présente trois essais sur la manière avec laquelle les éléments clés de l'environnement économique façonnent la dynamique des entreprises dans les pays en développement, notamment l'infrastructure, les frictions d'ajustement et les réglementations.

Dans le premier chapitre, Matias Busso et moi étudions les déterminants des effets agrégés et régionaux des nouvelles infrastructures de transport. Un canal clé négligé est le rôle que la politique en matière d'infrastructure joue dans le changement des incitatifs des entreprises à entrer, sortir et croître - générant ainsi des changements endogènes dans la productivité locale. Le chapitre documente et quantifie l'importance de ce méchanisme en utilisant des microdonnées mexicaines détaillées et un modèle d'équilibre général spatial incorporant la dynamique des entreprises. En exploitant les retards aléatoires dans la construction des autoroutes, nous montrons empiriquement que la productivité croît dans les villes dotés d'une meilleure infrastructure de transport. Les entreprises jouent un rôle critique dans la réalisation de ces résultats: les autoroutes augmentent la taille des entreprises, les taux d'entrée, les taux de survie et la productivité totale des facteurs. Ensuite, en calibrant notre modèle sur les données du recensement entre 1998 et 2018, nous constatons que les nouvelles autoroutes au cours de cette période ont augmenté le bien-être et le revenu d'un demi pourcent, ce qui est similaire à leurs coûts en termes de PIB. De plus, nous constatons une importante réaffectation spatiale des travailleurs et de la production. Près de la moitié de ces effets s'expliquent par des changements endogènes dans la productivité locale, entraînés par la dynamique des entreprises.

Dans le deuxième chapitre, co-écrit avec Jonas Gathen, nous nous concentrons sur les moteurs des miracles de croissance. Nous soutenons que les miracles de croissance sont stimulés par une course fondamentale: alors que l'économie essaie de rattraper son état stationnaire, les changements dans l'environnement économique déplacent l'état stationnaire lui-même et fournissent un nouveau potentiel de rattrapage de la croissance. Nous quantifions cette course au cours du développement en utilisant 40 ans de données de panel des entreprises manufacturières en Indonésie et un modèle structurel de dynamique des entreprises. Nous estimons le modèle sur les microdonnées le long du chemin de croissance observé, sans jamais supposer que l'économie ne soit à un état stationnaire. Alors que la croissance de rattrapage à partir des conditions initiales en 1975 représente 42% de l'industrialisation ultérieure de l'Indonésie, de nouveaux changements dans l'économie induisent de nouvelles croissances de rattrapage. En fin de compte, l'économie est dans une course sans fin où elle ne parvient jamais à rattraper son plein potentiel.

Dans le troisième chapitre, Santiago Levy et moi étudions un paradoxe de croissance courant dans les pays en développement, où une industrialisation rapide pourrait être couplée à une faible croissance de la productivité globale. Nous soutenons que le paradoxe peut s'expliquer par deux forces contradictoires. D'une part, les gouvernements introduisent des politiques qui favorisent la croissance, telles que la libéralisation des échanges, les agences de concurrence et les organes de réglementation ; et d'autre part, les politiques qui segmentent l'économie en secteurs formels et informels. Pour éclairer ce résultat, nous construisons un ensemble de données de panel sur 20 ans au niveau des établissements pour le Mexique, un pays où les exportations manufacturières sont passées de sept à 33% du PIB, mais où l'informalité du travail a à peine changé, l'informalité des entreprises a augmenté et la croissance de la productivité totales des facteurs était négative. Nous constatons que de nombreuses entreprises formelles à haute productivité sortent; la taille et la productivité des entreprises survivantes augmentent à peine, et beaucoup deviennent informelles. Les entrants sont moins productifs que les survivants, principalement en raison de l'entrée massive dans les secteur informel. Enfin, nous montrons que même si le secteur manufacturièr se comporte mieux, sa contribution à la TFP est modeste car l'informalité persiste dans le secteur. Malgré une croissance spectaculaire des exportations, le pays se désindustrialise maintenant.

Contents

List of Figures

List of Tables

Chapter 1

Building Up Local Productivity: Infrastructure and Firm Dynamics in Mexico

Matias Busso[1] & Oscar Fentanes [2]

Abstract

What determines the aggregate and distributional effects of new transportation infrastructure? One key overlooked channel is the role that infrastructure policy plays in changing the incentives of firms to enter, exit, and grow – in turn generating endogenous changes in local productivity. In this paper, we document and quantify the importance of this channel by using detailed Mexican microdata and a spatial general-equilibrium model that incorporates firm dynamics. Leveraging random delays in the construction of highways, we empirically show that productivity grows in places with better transportation infrastructure. Firms play a critical role in driving this results: highways increase firms' size, entry rates, survival rates, and total factor productivity. Then, by calibrating our model on census data between 1998 and 2018, we find that new highways over this period increased welfare and income by half a percent, similar to its costs in terms of GDP. Moreover, we find substantial spatial reallocation of workers and production. Nearly half of these effects are explained by endogenous changes in local productivity, which is driven by firm dynamics.

JEL Codes: D22, L25, O47, O54, R41, R42

Keywords: Firm dynamics, aggregate productivity, infrastructure

---

[1]Inter-American Development Bank; mbusso@iadb.org .

[2]Toulouse School of Economics; oscar.fentanes@gmail.com. For feedback and comments, we thank Fabrice Collard, Christian Hellwig, Patrick Fève, Eugenia Gonzalez-Aguado and Stéphane Straub; as well as participants of the TSE Macro workshop, MYE Turin, CIREQ Montréal and World Bank conferences. We especially thank Jonas Gathen, Chanwoo Kim, and Sébastien Montpetit for many discussions and suggestions. We thank Kyunglin Park and Mushegh Mkrtchyan for their amazing research assistance and Salvador Navarro and David Rivers for sharing the STATA code for productivity estimation. We thank Natalia Volkow and the Mexican Institute of Statistics and Geography for facilitating access to confidential microdata. All results have been reviewed to ensure that no confidential information is disclosed. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the the Inter-American Development Bank, its Board of Directors, or the countries they represent.

1.1 Introduction

Transportation infrastructure is a key determinant of economic development because it reduces trade costs and travel times for moving both goods and people, bolstering GDP and welfare (Banerjee et al., 2020; Allen and Arkolakis, 2022). Over the past years, economic geography models –the workhorse spatial framework to study transportation infrastructure– have emphasized the importance of locations' characteristics to understand the aggregate and distributional effects of such policies (Allen and Arkolakis, 2014; Redding and Turner, 2015). In this literature, locations are characterized by two fundamental features: 1) amenities, which include housing, weather, cultural attractions, and personal connections; and 2) local productivity, explaining why the same worker may be more productive in one place than in another. While the literature has shed light on key components of local amenities, such as housing supply or public goods congestion, it remains a challenge to understand what determines local productivity, which is often viewed as an exogenous feature mostly subject to agglomeration forces.

In this paper, we argue that local productivity is shaped by firm dynamics – that is, the endogenous processes of entry, exit, and growth; and, moreover, that such firm dynamics are a key driver of the effects of new transportation infrastructure. We support this idea by answering the following two research questions: Does new transportation infrastructure affect firm dynamics and local productivity? And, to what extent do firm dynamics drive the aggregate and distributional effects of infrastructure policy? We tackle the first question empirically by using firm-level panel data from Mexico and leveraging a natural experiment arising from highway planning and execution nationwide over two decades. We answer the second question quantitatively by proposing an economic-geography model à la Allen and Arkolakis (2014), extended with firm dynamics in the spirit of Melitz (2003).

Our central empirical result is that improvements in transportation infrastructure do, in fact, lead to local productivity growth, and that this increase is connected to changes in firm dynamics. These findings are based on two main data sources. First, the Mexican Economic Census, a detailed panel data set covering the universe of firms across all locations in the country. Second, the National Highways Network, a comprehensive digitization of all paved roads in Mexico – allowing us to fully characterize the dynamics of firms and the evolution of transportation infrastructure over a 20-year period, from 1998 to 2018.

The main empirical challenge is reverse causality, a concern because it is plausible that economic outcomes determine where the government chooses to build new highways. To overcome this issue, we implement a *delayed planned construction* approach by digitizing the placement and characteristics of 250 highways that were planned over the period from 2007 to 2018. In Mexico, presidents present their national highway construction plans when they begin their term, and they provide Congress with detailed progress reports throughout their tenure. We use these reports to track the execution status of the plans and their exact construction timing. The identifying assumption is that, while placement of construction plans may be influenced by demographic, political, and economic factors, the timing of actual execution, conditional on its previous selection, is as good as random.

Leveraging this source of variation, we estimate a staggered differences-in-differences model following Callaway and Sant'Anna (2021) for two sets of construction plans between 2007 and

2018. We categorize a firm as *treated* if it operates in a location close to an executed construction plan and as *not yet treated* if it is close to a plan that was not executed. We exclude from the sample all firms far from construction plans. Although this treatment is binary, we show that it implies significant increases in market access for treated locations.

According to our baseline point estimates, during the treatment period, workers in treated locations increase their labor productivity by around 5%. This can be explained by two mechanisms. First, firms in treated locations are themselves 2% more productive. Second, these firms also become 2% larger. Thus, local labor productivity grows because more workers are employed by more productive firms. Firms also become more likely to survive, generating persistence in the productivity composition of firms in treated locations. After five years, both workers and firms still exhibit higher productivity in treated locations, indicating long-lasting effects of highways. Moreover, these effects are accompanied by higher entry rates, suggesting that potential entrants also react to new transport infrastructure but take longer to respond.

We use a unified framework of economic geography with firm dynamics to theoretically decompose the benefits derived from improvements in transport infrastructure into two parts: gains resulting from reduced trade costs and gains stemming from local productivity growth driven by firm dynamics.

As is standard in static economic-geography models, our model features a country with a large number of locations (e.g., cities or municipalities) that differ in exogenous amenities and local labor productivity. These two characteristics, combined with the geography of trade costs, determine the spatial distribution of workers, wages, and outputs. The innovation of our model is that local productivity is given by the average of firms' productivities. Thus, local productivity is essentially determined by the number and composition of firms. Because firms' decisions about entry and exit are endogenous and dynamic, so is local productivity.

Our model highlights an important mechanism linking transport infrastructure, firm behavior, and local productivity. Suppose that the government builds a new highway to connect two important cities. Firms in locations along the road's path will benefit from greater market access. They will face lower trade costs, allowing them to sell their products in more distant markets and to lower prices for their goods. This boosts incumbent firms' size and profits, and therefore, their survival probability. Potential entrants observe the higher profitability of active firms, thus increasing the likelihood that new firms indeed enter. Crucially, the more productive and larger the firm, the higher its probability of entering and surviving. Thus, although the increase in market access benefits all firms regardless of their productivity, it reinforces the entry and persistence of large and productive firms. As a consequence, the productivity of locations along the new highway increases. The opposite is also the case; that is, the productivity at locations not connected by the new highway stagnates or decreases.

We recover model fundamentals through a sequential combination of parameterization, model inversion, and internal calibration to match the path of spatial equilibria in the economic census from 1998 to 2018. We calculate the geography of trade costs by computing the minimum travel times between any pair of locations and parametrically mapping them to iceberg costs. We determine the path of amenities and labor productivity by inverting the system of spatial equilibrium equations. Intuitively, differences in population identify differences in amenities and differences in wages identify differences in local labor productivity. Finally, we recover the parameters governing the distribution of firm-level productivity and the entry and exit processes through internal calibration. The local firm size distribution identifies the productivity distribution, and local entry and exit rates identify entry and exit costs.

The calibrated model shows that new highways in Mexico from 1998 to 2018 contributed to real income and welfare growth, and that these benefits were unequally distributed. It reveals that firm dynamics played a central role in these effects. In line with the conclusions of previous, static studies (Allen and Arkolakis, 2014), our findings document that new highways increased welfare by 0.44% and increased aggregate real revenues by 0.64% in 2018; since Mexico invests annually 0.5% of its GDP in transport infrastructure, our results suggests that the policy might be cost effective.

These aggregate effects hide substantial distributional effects. The areas that experienced the largest investments in new transportation infrastructure were in three key locations: those near the California and Texas borders, those close to the major ports serving Asia and Europe, and those close to the Caribbean Sea. In these locations, infrastructure improvements significantly reduced trade costs and improved market access, enhancing the relative competitiveness of firms that could tap these benefits. As a result, both real revenues and populations in these areas increased by nearly 10%, largely at the expense of the central regions of the country that were largely bypassed by highway infrastructure investments.

To understand how firm dynamics contributes to welfare and real income gains, we compare our baseline results to those from a model without firm dynamics – that is, one in which local productivity is exogenous and policy invariant. We find that productivity gains driven by firm dynamics explain up to 46% of the overall real income gains, and that the rest of the gains stem from reductions in trade costs. Moreover, in a model without firm dynamics, the distribution of income gains is more uniform, suggesting that firm dynamics are a force for spatial divergence.

Finally, we find that productivity grains are mostly driven by better firm selection. This finding comes from decomposing local productivity gains due to highways into two endogenous components: firm selection, as measured by average idiosyncratic firm productivity, and the number of firms. We find that the firm selection accounts for 77% of the productivity gains, and the increasing number of firms explains the remaining 33%.

Overall, our quantitative results show that new highways in Mexico had a more significant impact on the spatial reallocation of economic activity than on aggregate welfare and income. This finding conveys an important message to policymakers: transportation infrastructure can serve as a powerful tool for shaping the geographical distribution of economic activity by providing incentives for workers and firms to operate in specific locations.

Related literature and contributions

Our contribution is twofold. First, we offer new evidence on the effects of infrastructure on firm dynamics using panel data for all economic units in a developing country. Second, we develop a spatial general equilibrium framework where endogenous firm dynamics determine local productivity. In doing so, we establish a bridge between empirical research on the effects of infrastructure on firms and the dynamic spatial literature that quantifies the aggregate and distributional effects of place-based policies.

Economic geography. This paper builds on the work of Allen and Arkolakis (2014); Redding (2016); Allen and Arkolakis (2022). We extend their framework by incorporating firm dynamics. This approach endogenizes local productivity, allowing us to decompose the income and welfare gains resulting from new transportation infrastructure into two component parts: the reductions in trade costs, and the growth of local productivity.

Dynamic spatial models: Our paper relates to recent dynamic spatial frameworks. Using an approach similar to that of Caliendo et al. (2019), we present a model with trade and labor mobility; however, we allow for firm heterogeneity in a non-competitive market. Similar to the work of Lindenlaub et al. (2022), we focus on firms; using an approach similar to the one adopted by Kleinman et al. (2023), we also feature a dynamic spatial trade model with labor mobility. However, there are important differences in that Lindenlaub et al. (2022) abstract from the trade structure, and Kleinman et al. (2023) assume a representative firm by location with exogenous productivity. In contrast to both, we allow local productivity to be determined by the dynamics of heterogeneous firms in an internal trade environment. To the best of our knowledge, this is the first paper to incorporate entry, exit, and growth dynamics of heterogeneous firms in a spatial model with a realistic geography of trade costs, and then to validate it with a natural experiment.

Effects of infrastructure on growth. This paper also relates to the micro-empirical literature that measures the effects of transport infrastructure on local growth (Donaldson, 2018; Banerjee et al., 2020) and firm performance (Holl, 2016; Holl and Mariotti, 2018; Gibbons et al., 2019). Our contribution lies in providing new evidence for a developing country by using novel firm-level panel data that cover the universe of firms from all industries, both formal and informal, over a 20-year period. To the best of our knowledge, this is the first paper in this literature that features data of such comprehensive coverage for a developing country.

Furthermore, previous studies focusing on the effects of transport infrastructure on firm-level productivity have relied on traditional estimation procedures such as those used by Levinsohn and Petrin (2003) and Olley and Pakes (1992). However, these measures are based on value-added production functions and confound the effects of infrastructure on revenues and intermediate inputs. Our paper estimates firm productivity using a gross output-production function similar to that of Gandhi et al. (2020). This approach reveals productivity gains stemming only from higher revenues, in line with standard trade models.

Effects of infrastructure on firm dynamics. Evidence on the effects of infrastructure on firm dynamics is scarce because of data limitations. Among these few studies, Shiferaw et al. (2015) document that better transportation infrastructure favors firm entry, especially of large firms. Zhou (2023) also finds that locations with better exposure attract larger firms, but that places far from highways have higher entry rates. Our paper provides new evidence on entry and exit by documenting that responses to the arrival of new infrastructure are faster for exits than for entries. Moreover, our paper is the first one to show that while infrastructure does induce within-city firm migration, the entry and exit impacts of infrastructure are not the result of firms migrating across cities.

Effects of highways in the Mexican context. Other empirical studies have focused on the impacts of road infrastructure in Mexico. Examples include Durán-Fernández and Santos (2014); Pérez and Sandoval (2017); Blankespoor et al. (2017). These studies have relied on location-level data. By contrast, by exploiting panel identifiers from Busso et al. (2018), we show firm-level time variation for the first time. Moreover, our study examines the impacts of a more detailed and denser highway network, and it exploits execution of presidential construction plans as a source of exogenous variation to measure causal effects.

Structure of the paper. The rest of the paper proceeds as follows: Section 1.2 briefly discusses the economic and infrastructural context of Mexico. Section 1.3 discusses the sources, novelty and advantages of our data. Section 1.4 outlines our empirical approach and presents our results. Section 1.5 shows our dynamic spatial general- equilibrium model. Section 1.6 shows how we estimate the model and how the model fits the data. Section 1.7 presents our quantitative results. Section 1.8 concludes.

1.2 Growth and infrastructure in Mexico

After the implementation of macroeconomic policies inspired by the Washington Consensus in the 1990s and the North American Free Trade Agreement (NAFTA) in 1994, Mexico has enjoyed an extended period of macroeconomic stability (Levy, 2018). Nevertheless, in terms of real GDP, the nation has seen an average annual growth rate of merely 2.4% between 1995 and 2015, resulting in a corresponding annual growth of real GDP per capita of just 0.8%.

Economic growth has been not only slow but also unequally distributed across regions. Between 1995 and 2015, states near the US border, such as Chihuahua and Nuevo León, or in the central industrial belt such as Guanajuato and Querétaro, experienced rapid industrialization, resulting in annual real GDP growth rates exceeding 4%. Conversely, states in the southern region, such as Chiapas, Guerrero, and Oaxaca, remained largely underdeveloped and achieved an average real GDP growth rate of a mere 1% real over the same period.

A prevalent explanation for these disparities in economic performance is the unequal distribution of high-quality transport infrastructure. Regions with limited access to highways, railroads, and seaports are less appealing to firms that rely on high connectivity to intricate input-output networks (Dávila et al., 2002). In this perspective, highways are of unrivaled importance for Mexico, given that 83% of domestic cargo is transported via road freight.[3]

For the past two decades, the federal government has acknowledged deficiencies in the highways network and sought to address them through the sexennial National Infrastructure Plan. In these plans, the government determines the objectives, location, characteristics and budget of key, proposed highways. However, while most middle-income countries allocate between 1% and 5% of annual GDP to new, inland transportation infrastructure, Mexico's investment is only around 0.5% of GDP (OECD, 2020). As a result, Mexico's investments almost certainly insufficient to meet its transportation needs.

The extent to which this deficiency in robust transportation infrastructure might contribute to the country's sluggish economic growth, despite the implementation of ambitious macroeconomic reforms, continues to be a subject of ongoing debate. Moreover, it remains an open

[3] In addition, 96% of people traveling within the country use highways.

question as to whether a policy of more ambitious investments in road infrastructure in underdeveloped regions could potentially attract highly productive firms and reduce economic disparities across the country.

1.3 Data

Our study relies on three primary sources of data. The first is the Mexican Economic Census, which is conducted every five years. We rely on data collected from 1998 to 2018. These data have three important features: they cover the universe of establishments in Mexico, provide geolocations of establishments at the block level, and they longitudinally link establishments. These features allow us to characterize firms' dynamics across all locations. A second key data source is the National Highways Network. We use data from the network over the period from 2004 to 2019. These data allow us to determine all origin-destination travel times and to estimate trade costs between locations. A third key source of data is the National Infrastructure Plans from the presidential terms over the periods from 2007 to 2012 and 2013 to 2018. These plans describe how each new administration intends to spend its infrastructure budget. Here we provide next a brief overview of each data set's characteristics. (Greater details on data construction and cleaning procedures are provided in Appendix 1.A.)

1.3.1 The Economic Census

Our main data source is the Mexican Economic Census, collected by the Mexican Institute of Statistics and Geography (INEGI). Although the census is conducted at the establishment level, throughout our paper we refer to these units as firms.[4] The census captures all formal and informal establishments of all sizes that produce goods or provide services in fixed facilities. The census includes such facilities in all locations with a population larger than 2,500 people and for all 6-digit industries according to the North American Industrial Classification System (NAICS). Excluded from the census are agriculture and government (and street vendors of any industry). In this paper, we focus on establishments in manufacturing, commerce, and service sectors. To leverage the panel structure of the census, we use INEGI's official firm identifiers to link the waves in 2008, 2013, and 2018. To link the waves 1998, 2003, and 2008, we use the fuzzy linkage described in Busso, Fentanes and Levy (2018), which uses firm identity, location, and industry to match units across census waves.[5]

Table 1.1 displays the coverage of the census, indicating that the number of firms increased from 2.7 million in 1998 to 4.7 million in 2018, representing an implied annual growth rate of 2.8%. Over the same period, the number of workers increased from 13.3 million to 24.8 million, with an implied annual growth rate of 3.1%. For reference, the corresponding average GDP growth rate was 2.4%.

Based on annual employment surveys, there were an estimated 39 million workers in urban the locations that were included in the 2018 economic census, (i.e., places with more than 2,500 people). Table 1.1 reveals that our data encompass almost 25 million workers, representing 61.5% of the national workforce. The difference between these figures is due to the government

[4] Levy (2018) documents that 99.7% of establishments are single-establishment firms.

[5] The accuracy rate of this linkage algorithm is 95% (Busso et al., 2018).

Table 1.1: Mexico's Economic Census

Year	Firms (millions)	Workers (millions)	6-digit sectors	Populated locations
1998	2.72	13.31	720	2,566
2003	2.92	14.41	726	2,629
2008	3.66	18.14	732	2,801
2013	4.17	19.66	735	3,033
2018	4.73	24.82	741	3,234

Notes: Full census coverage.

sector, which employs 4 million workers, and the remaining 10 million workers who operate as street vendors.[6]

Locations. The Economic Census stratifies the territory into three primary levels: state, municipality, and locality. While the boundaries and codes for states and municipality remain constant, those for localities may change because they are based on demographic characteristics that may require redefining a given census tract's boundaries. To account for differences in census tracts, we establish our own fixed geography. We accomplish this by defining a time-consistent set of locations, composed of localities likely to belong to the same city. The procedure consists on generating a 1 km buffer around the 7,136 localities and classifying contiguous buffers as the same location. This procedure results in 3,248 locations consistent across all census waves. Panel (a) in Figure 1.1 shows their geographic distribution.[7]

1.3.2 The National Highways Network

The second data source is the National Highways Network (*Red Nacional de Carreteras*). This database, published by INEGI, consists of shapefiles including all national and state paved roads and highways in Mexico at five points in time: 2004, 2011, 2014, 2018, and 2019. Panel (b) in Figure 1.1 illustrates this network in 2018. In 2004, Mexico had 106,079 kilometers of paved highways, by 2019, the network reached 187,453 kilometers. [8]

We use the data on highways to create a matrix of minimum travel times between any two locations in the country to help estimate internal trade costs for our quantitative model. With 3,248 locations defined, the size of our minimum-travel-times matrix is 3,248$\times$3,248. To compute it, we implement the Dijkstra (1959) algorithm, which finds the shortest path between two nodes in a network. We reduce the digitization bias pointed out by Allen and Arkolakis (2014), by discretizing the space into a grid of 382,181 hexagons.[9] Each hexagon is weighted by the maximum legal speed on the highways that cross them. If two or more highways cross a hexagon, we use only the highway with the top maximum speed. If a hexagon belongs to the interior of a city, we assume that its speed is 30km/h. Hexagons carry information about how

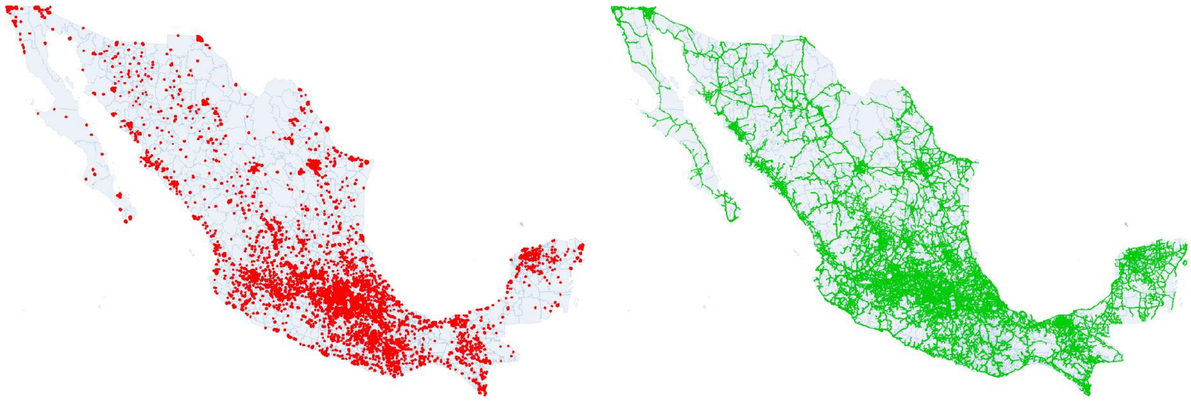
[6]Table 1.1 also shows that the number of 6-digit sectors slightly increased from 720 6-digit industries in 1998 to 741 in 2018. This is mostly due to revisions of the NAICS.

[7]Table 1.1 shows that the economic census increases its geographic coverage over time. The main reason is that, as the population grows, more localities cross the 2,500-person threshold, and thus they qualify to appear in the economic census.

[8]The comparable network in France is close to 1 million kilometers (*Autoroutes nationales, départementales et communales*). To put this into perspective, France at that time had 14 meters of highway per capita, 10 times the 1.4 meters per capita figure for Mexico..

[9]The edge length is 1.22 kilometers. The H-resolution is 7 according to Uber's Hexagonal Hierarchical Spatial Index.

Figure 1.1: Locations and highways network in Mexico, 2018

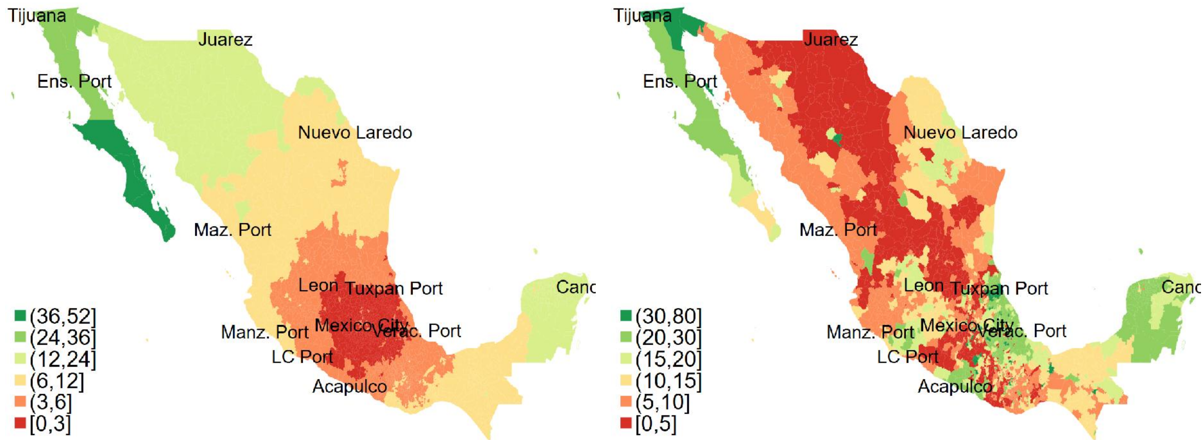


(a) 3,248 locations (b) 187,000 km of highways

Notes: Panel (a) shows locations following our definition. Panel (b) all paved roads and highways excluding within-city roads.

level or steep the terrain is; this is considered in the computation. (Appendix 1.A provides additional details.)

Figure 1.2: Estimated minimum driving times to Mexico City



(a) Estimated minimum number of hours needed to drive from municipalities to Mexico City (2018)

(b) % decrease in estimated minimum number of hours to drive from municipalities to Mexico City (1998-2018)

Notes: Maps subdivided into municipalities.

Panel (a) of Figure 1.2 shows the minimum travel times required to drive to Mexico City from the 2,457 other municipalities in the country. Assuming no traffic jams, 70% of municipalities can be reached from Mexico City within 6 hours; 20% take between 6 hours and half a day, and the remaining 10% require at least half a day. The most remote location is a 52-hour drive from Mexico City.

Panel (b) of Figure 1.2 illustrates the percentage change in time required to reach Mexico City during the period 1998-2018. The time needed to reach Mexico by road decreased by less than 10% over that period. For nearly one-third of Mexico's municipalities the time needed to reach the capital declined by 10% and 20%. For roughly one-fifth of the municipalities, the time needed to drive to the capital decreased by more than 20%. The regions that saw the most significant improvements, shown in green on the map, include those near the Caribbean Sea, the California and Texas ports of entry, and the two primary seaports connecting the country to Europe and Asia.

1.3.3 The National Infrastructure Plans

Our third data source is the National Infrastructure Plans. These data contain 250 construction plans from 2007 to 2018. They provide a source of quasi-natural variation that we utilize in our empirical analysis. The plans originate from two distinct presidential terms: 175 from the Felipe Calderón administration (2007-2012) and 75 from the Enrique Peña Nieto administration (2013-2018).

We use geographical software to locate all 250 plans on a map. If plans are executed, their locations can be easily pinpointed on a map since they appear in subsequent waves of the highway network shapefiles with updated characteristics. However, in cases where the plans are not executed, we infer their locations based on the plan descriptions. Subsequently, we draw these hypothetical highways on our shapefiles and assign them attributes such as width, number of lanes, and maximum speed based on the technical specifications provided in the construction plans.

To accurately document plans' execution and timing, we relied on annual progress reports from the Mexican Transportation Ministry to Congress. These reports provide detailed information on the number of kilometers built each year, the amount of money spent, and the year of project completion. It is important to note that highway construction plans may or may not be executed for various reasons. The actual execution of a project could be influenced by budgetary changes, technical challenges, opposition from the local population, or other political considerations. Unfortunately, the reports do not specify the reasons why a given plan was not built. In the empirical section, we examine whether plan execution and timing can be predicted by the characteristics of adjacent cities.

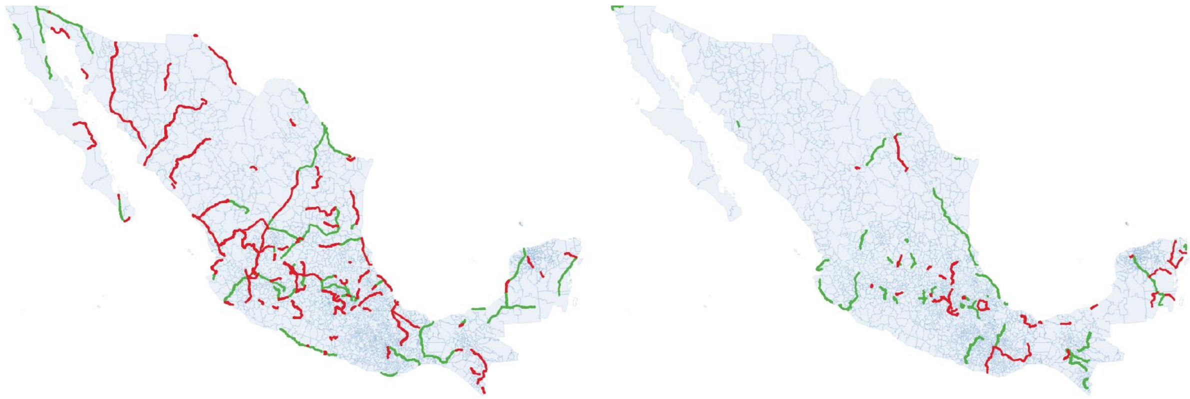
Table 1.2: Construction plans and year of execution

(a) 2007-2012 Administration			**(b) 2013-2018 Administration**		
Executed	Execution year	Total	Executed	Execution year	Total
No		115	No		33
Yes	2007	2	Yes		
	2008	11		2013	9
	2009	9		2014	10
	2010	7		2015	4
	2011	10		2016	9
	2012	21		2017	10
Total		175	Total		75

Table 1.2 presents the execution status of the construction plans and their timing. For the administration over the period from 2007 to 2012, 40% of the 175 construction plans were fully executed. Half of these plans were completed within the first four years of the presidential term, while the remaining half were finished in the last two years. Similarly, for the administration over the period from 2013 to 2018, 56% of the 75 construction plans were completed, with half of them being finished in the first 2 years.

Figure 1.3 shows the geographical distribution of plans according to their execution status. The

Figure 1.3: The construction plans.

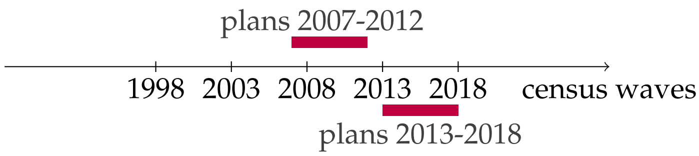


(a) Plans 2007-2012 (b) Plans 2013-2018

Notes: Green lines denote construction plans that were completed. Red lines denotes plans that were not built.

majority of states were crossed by at least one plan during the first administration. In contrast, most plans during the second administration were concentrated in the southern region of the country.

Figure 1.4: Overlap construction plans and Economic Census waves



We combine these construction plans with economic census data. As shown in Figure 1.4, there is no perfect temporal overlap between the two databases. We leverage this fact to characterize pre-treatment, treatment, and post-treatment periods. Clearly, the censuses in 1998 and 2003 serve as pre-treatment periods for both sets of construction plans, and the censuses from 2008 to 2018 serve as staggered treatment periods.

1.4 Empirical evidence

We document how improvements in transportation infrastructure affect local labor productivity and firm dynamics. We first combine data from the Mexican Economic Census with information from the National Highways Network, and then leverage the timing of the execution of presidential construction plans as a source of plausibly exogenous variation to estimate a staggered differences-in-differences regression model, as in Callaway and Sant'Anna (2021).

Measuring the effects of infrastructure on economic outcomes is a challenging task for two main reasons. First, the placement of infrastructure projects is not random; most of the time, economic or political considerations motivate the placement. Second, infrastructure projects, such as highways, may produce spillover effects because such projects are part of a larger network that can benefit all locations to varying degrees. Our empirical approach explicitly addresses the first problem. However, in our baseline specification, we do not account for spillover effects. If spillovers do exist, our estimates would represent a lower bound.[10]

1.4.1 Main specification

The main goal of our empirical section is to document that when firms are exposed to better transport infrastructure, their performance improves, and so do their chances of entering and surviving. Moreover, we are interested in disentangling whether these effects stem from the actual construction or simply the announcement of new highways. Finally, we aim at determining if highways have only a temporary effect on firms or if they persist in the medium-run. To capture these effects we rely on a standard differences-in-differences model with staggered treatment timing. The regression equation is of the form:

$$y_{n,i,t} = \alpha_i + \alpha_t + \gamma' \mathbf{X}_{n,i} + \sum_{e=1}^{min} \beta_e^{persist.} \cdot D_{i,t-e} + \beta_0 \cdot D_{i,t} + \sum_{e=1}^{max} \beta_e^{anticip} \cdot D_{i,t+e} + \varepsilon_{n,i,t} \quad (1.1)$$

In (1.1), the index n denotes the firm, i denotes the location, and t denotes the period. On the left-hand side, $y_{n,i,t}$ represents the outcome of interest. On the right-hand side, the coefficient α_i indicates location and α_t corresponds to time fixed effects. $\mathbf{X}_{n,i}$ is a vector of observed control variables. The error term ε_{it} is clustered at the location level, at which the treatment occurs as is standard in the literature.

The treatment variable, D_{it}, is defined at the location level i. It is a binary indicator that takes the value of one for firm n if its location is exposed to the execution of a construction plan at time t and zero otherwise. The model includes three conceptually different treatment effects. The set of coefficients $\beta_e^{persist.}$ captures the effects of the treatment before period t on current outcomes. The coefficient β_0 measures the contemporaneous treatment effect at t. Finally, the set of coefficients $\beta_e^{anticip.}$ reflects possible anticipatory effects at t of future treatments. We provide a detailed description of D_{it} below.

For the coefficients $\beta_e^{persist.}, \beta_0, \beta_e^{anticip.}$ to be identified, the model relies on the following assumptions. The first concerns the *irreversibility of the treatment*; that is, once a highway is built, it cannot be destroyed. This assumptions ensures that the three groups of coefficients are separately identified. The second concerns *conditional parallel trends* based on a never-treated group; that is, only firms in a location with the same characteristics would follow the same trend in the absence of treatment. This assumption guarantees that the measured effects can have a causal interpretation. Following Theorem 1 in Callaway and Sant'Anna (2021), these assumptions imply that we can identify all group-time average treatment effects (ATE).

1.4.2 Outcomes

In this section we provide a detailed definition of our outcomes of interest: value added per worker, firm-level total factor productivity (TFP), firm size, and firm entry and exit rates.

Value added per worker. We calculate this by dividing firm value added by the number of workers. Value added is defined as the total revenue derived from all commercial activities of the firm, minus intermediate expenditures, such as raw materials and electricity. The definition

[10]We are working on a robustness check to account for spillover effects by specifying counterfactual infrastructure shocks as in Borusyak and Hull (2020).

of total workers includes blue- and white-collar employees, as well as owners, outsourced personnel, and piece-rate workers. We represent this metric logarithmically as $log(VA/L)$. This measure offers the advantage of being consistent with models that use standard frameworks in which the production function is constant returns to scale, and it relies solely on labor.

Revenue productivity. We measure TFP as in Gandhi et al. (2020) (henceforth GNR). This measure assumes a Cobb-Douglas production function of the form $y = TFP \cdot k_s^{\alpha} l_s^{\beta} m_s^{\gamma}$, where y represents gross output, k is the capital stock, l is total workers and m is the intermediate inputs. The three input elasticities, $\alpha_s, \beta_s, \gamma_s$, are assumed to be the same for all firms within the same three-digit industry s. We express this outcome logarithmically as $log(TFP)_{GNR}$. The main advantage of this productivity measure is that it attributes all increases in TFP to higher revenues while holding inputs constant. Traditional value added-based production functions such as Olley and Pakes (1992); Levinsohn and Petrin (2003); Ackerberg et al. (2015) cannot disentangle whether an increase in TFP is due to higher revenue or reductions in intermediate input expenditures. It is important to note that $log(TFP)_{GNR}$ is a revenue productivity measure. This means that it cannot disentangle whether a higher TFP is due to an increase in prices or an increase in physical productivity. This issue can be solved by exploiting firm-level prices; unfortunately, such data are unavailable in the Economic Census.

Firm size. This is simply the sum of all blue- and white-collar workers, owners, and outsourced and piece-rate workers. We denote this outcome in logs as $log(L)$. We consider owners and family members as part of production workers since most firms in Mexico operate exclusively with type of workers in profit-sharing agreements.

Average wage. This is measured as the total wage bill divided by the number of workers. We express this outcome in logarithms as $log(w)$. In cases in which firms do not report the wage bill because they operate under profit-sharing agreements that are commonly used by most informal firms we employ the wage imputation method described in Busso et al. (2012). This procedure involves assigning missing wages to be the same as those in firms from the same state, six-digit industry, and of similar size.

Entry and exit. For entry, this is a dummy variable that takes the value of one if the firm appeared for the first time in the census wave, and zero otherwise. For exit, it takes the value of one if the firm is observed for the last time and zero otherwise.

1.4.3 Treatment and sample

Treatment. The treatment variable, denoted as $D_{i,t}$, is an index function that equals one if the firm operates close to a fully executed construction plan and zero otherwise. A location is considered to be close to a construction plan if it overlaps with a buffer of radius B around the plan. For robustness checks, we consider different values for B, specifically, 5, 10 and 15 kilometers.

Figure 1.17 in Appendix 1.B illustrates treated (in green) and not-yet-treated (in red) locations for a specific buffer size B. Notice that the treatment is not defined for locations that do not overlap with any buffer; this will affect the sample size.

Sample. The sample includes only locations overlapping with construction plans. Table 1.3 shows the number of locations in the sample for $B = 5$ kilometers. It shows that 771 of the

3,248 locations overlap with construction plans from 2007 to 2012. Among them, 259 intersect with plans that were fully executed before 2012, and another 512 intersect with plans that were not undertaken. Similarly, for the construction plans envisioned over the period from 2013 to 2018, 457 locations overlap with construction plans; among them, 278 were fully executed before 2018. Table 1.14 in the appendix shows how the number of locations in the sample increases when we use a larger buffer size.

Table 1.3: Locations in sample and treatment group

Plans period	**2007-2012**	**2013-2018**
With plans	771	457
With out plans	2,475	2,789
Total locations	**3,246**	**3,246**
Executed	259	278
Not executed	512	179
Total locations	**771**	**457**

Although the Economic Census covers from 2.7 million firms in 1998 to 4.7 million in 2018 (see Table 1.1), we do not include all of them in our empirical estimation. Our sample is limited to firms in locations overlapping with construction plans. For instance, considering the construction plans from 2013 to 2018, Table 1.4 shows that 2.73 million firms in 2018 are in the sample. Among them, 1.26 million are in the treatment group. Table 1.15 in the appendix shows how the number of firms in our sample increases as we increase the buffer size.

Table 1.4: Firms in the sample and treated group

Plans period	**2007-2012**		**2013-2018**	
Census	Sample	Treated	Sample	Treated
1998	2.09	1.43	1.65	0.73
2003	2.23	1.51	1.75	0.77
2008	2.72	1.82	2.14	0.97
2013	3.06	2.04	2.43	1.12
2018	3.43	2.26	2.73	1.26

Notes: Treated means that the firm belongs to the treatment group, not that it was treated at that period.

1.4.4 Validity

The validity of our empirical approach relies on the timing of execution of construction plans being orthogonal to economic outcomes. We provide three tests to show that this source of variation is indeed as good as random.

The first test evaluates whether execution of plans can be predicted. We show that while the geographical assignment of construction plans is correlated with demographic, economic, and political characteristics, the actual execution and the timing is not. To do this, we regress, at the location level, an index variable denoting if a location is close to a construction plan, and whether it was executed, on local characteristics. Column (1) from Table 1.16 (Appendix 1.B) shows that certain areas are more likely to be targeted by a construction plan. These areas are those that have larger populations and higher value added per worker, and those that voted for the opposition party in the previous presidential election. Column (2), however, shows that none of these characteristics matter for the eventual execution of the plan.

The second test provides a balance table to study whether treated and untreated locations differ in characteristics at baseline. We find that although there are some differences in levels, there are none in growth rates. Table 1.17 (Appendix 1.B) shows that treated and untreated groups are similar in population, average firm size, average firm productivity, and industrial composition. Firms in treated locations, however, hire more formal workers on average, and they are more capital intensive. If we focus on growth rates, they don't seem to evolve differently, which suggests that existing differences are constant across time.

The third test addresses whether the parallel-trends assumption holds. As we show in the following section (tables 1.5 and 1.6), there are no statistically significant pre-trends in our outcomes of interest.

Additional concerns about the validity of our approach are that construction plans may only capture minor improvements in the highways network, and that most of the effects that we measure may be driven by other infrastructure projects tied to the plans, such as industrial parks or housing developments. In Table 1.18 (Appendix 1.B) we provide evidence that the construction plans have a significant effect market access. Execution of construction plans imply an increase in market access of 0.07%. Because the baseline increase was on average 0.13%, the implied gains derived from plan execution are 53%.

1.4.5 Empirical results

To derive our baseline results we estimate two separate event-study models following Callaway and Sant'Anna (2021). One is for the 2007-2012 construction plans, and the other is for the 2013-2018 plans. Regressions are estimated at the firm level, assuming that the data are repeated cross-sectional.

Tables 1.5 and 1.6 show our baseline results for a buffer size $B = 5$km. As is good practice in the literature (Baker et al., 2022), we show first our estimates without covariates. Appendix 1.B shows robustness checks that include firm- and location-level covariates; we separately estimate the model for tradable and non tradable goods, and for different buffer sizes.

Table 1.5: Baseline results. Construction plans 2007-2012

	(1) $log(va/L)$	(2) $log(TFPR)_{GNR}$	(3) $log(L)$	(4) $log(w)$	(5) $Entry$	(6) $Exit$
β_{-1}	-0.0167	-0.0041	-0.0014	-0.0059	0.0083	0.0074
s.e.	[.0228]	[.0135]	[.0084]	[.0056]	[.0069]	[.0072]
β_0	-0.0017	0.0044	-.0139**	-0.0015	0.0143	-.0169**
s.e.	[.0158]	[.0056]	[.0063]	[.0057]	[.0101]	[.007]
β_1	.0653***	.0174**	-0.0049	-.0111**	.0158***	-.0329**
s.e.	[.0197]	[.0082]	[.0098]	[.0054]	[.0053]	[.0129]
Controls	No	No	No	No	No	No
Obs.	7,060,649	7,060,649	7,060,649	7,060,649	7,060,649	7,060,649

Notes: Standard errors are in brackets. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$. The sample includes all firms from 1998 to 2018. It excludes firms with value added or capital levels smaller than zero.

Construction plans 2007-2012. Table 1.5 shows that, for all outcomes of interest, β_{-1} is not statistically different from zero, which suggests that firms did not react to the treatment before they were exposed, and thus, that the parallel-trends assumption holds. According to the estimated β_0, there is no evidence of a contemporaneous effect of construction plan execution on labor productivity and firm TFP. There is a negative effect on firm size, but this does not have a significant effects on wages. Though there are no contemporaneous effects on firm entries, there are effects on plan exits; plan execution decreases firm exits by 1.69 percentage points. Finally, the coefficients β_1 capturing effects of highways five years after their construction, show a 6.5% increase in labor productivity and a 1.74% increase in firm TFP. Firm size is not affected. Wages slightly decrease. Firm entries increase by 1.6 percentage points, and firm exits decrease by 3.3 percentage points.

In summary, although the results for the 2007-2012 construction plans are in line with a story of labor and firm productivity gains and changes in firm dynamics due to better transport infrastructure, the results also suggest that these effects may take time to unfold.

Table 1.6: Baseline results. Construction plans 2013-2018

	(1) $log(va/L)$	(2) $log(TFPR)_{GNR}$	(3) $log(L)$	(4) $log(w)$	(5) $Entry$
β_{-1}	-.0641**	-0.0073	-0.0028	0.002	0.0086
s.e.	[.0257]	[.0075]	[.0076]	[.0048]	[.0055]
β_0	.0547**	.0179**	.0157*	.0113**	0.0018
s.e.	[.0226]	[.0074]	[.0092]	[.005]	[.0079]
β_1	-0.0013	0.0094	.0335***	.0146**	.0238*
s.e.	[.0231]	[.0112]	[.0109]	[.0065]	[.0124]
Controls	No	No	No	No	No
Obs.	6,375,668	6,375,668	6,375,668	6,375,668	6,375,668

Notes: Standard errors are in brackets. $^{***}p < 0.01,^{**} p < 0.05,^{*} p < 0.1$. The sample includes all firms from 2003 to 2018. The sample excludes firms with value added or capital levels smaller than zero.

Construction plans 2013-2018. Table 1.6 shows that there are no anticipatory effects for all outcomes of interest, except one, $log(va/L)$. The estimates for β_0 suggest significant contemporaneous effects on productivity and firm dynamics. When construction plans are executed, local labor productivity increases by 5.5%. This increase is coupled with an 1.8% rise in firm TFP and a 1.6% increase in firm size. There is a positive but noisy effect on average wages. There is no effect on firm entry.[11] One period after the treatment – that is, five years – neither workers nor firms in treated locations continue to be more productive. However, they become even larger (3.3%), and firm entry increases by 2.4 percentage points.

In summary, the results for 2013-2018 construction plans are in line with our hypothesis that better transportation infrastructure increase local labor productivity and that this is linked to more productive firms and changes in firm dynamics, notably, higher survival and entry rates. Although the effects on productivity are not persistent, effects on firm dynamics are.

1.4.6 Robustness checks

Regressions by sector. In Appendix 1.B we show the results estimated separately by three broad sectors: manufacturing, commerce and services. Table 1.19 shows the results for the

[11] Exit is not defined for this regression as the Census 2023 was not yet available at the time of this writing.

2007-2012 construction plans. The contemporaneous negative effects of highways on exit are present in all sectors although they are stronger in manufacturing and services (2 percentage points) than in commerce (1.4 percentage points). As in the aggregate case, the effects on productivity and firm dynamics appear one period after the treatment. The largest increases in labor productivity (8.4%) and firm TFP (3.6%) are in the commerce sector. Higher entry rates are present only in the services sector and lower exit in all. Overall, the sector that has the strongest response to new transportation infrastructure is commerce. This is not surprising since this is the sector for which trade costs are more relevant for input and output markets.

Table 1.20 shows the results for the 2013-2018 construction plans. Results by sector show that the contemporaneous positive effects on productivity are driven mostly by the commerce and services sectors and not by manufacturing. Again, there are no contemporaneous effects of firm entry. One period after the treatment, there are negative effects on revenue productivity for manufacturing firms, no effects for commerce, and positive effects for services. Higher entry five years after the treatment is only present in the commerce sector.

Controls. Tables 1.21 and 1.22 in Appendix 1.B show that our results are robust to adding time-invariant controls that take the Economic Census 1998 and the Population Census 2000 as baseline. For the 2007-2012 construction plans, point estimates preserve the sign, are slightly larger, and increase their statistical significance. A similar pattern is observed for the 2013-2018 construction plans.

Different buffer sizes. Tables 1.23, 1.25, 1.24 and 1.26 in Appendix 1.B show that our results are robust to larger buffer sizes. Focusing on the 2013-2018 construction plans, as the buffer size increases, most point estimates decrease and some become statistically zero. This is consistent with the fact that as we increase the buffer size, we also increase the sample and the risk of considering more distant locations as treated when they are only weakly affected by the construction plans.

Firm mobility. A common source of bias when studying firm entry and exit is firm relocation. When relocations are not tracked in the data, address changes are counted as an exit and then recounted as a new entry, biasing both rates upwards. Since we can track firm location changes across the entire country in the census, we verify whether new transport infrastructure incentivizes firms to relocate. In Appendix 1.B we show that although firm relocation can be substantial, as much as 5% of all surviving firms, they mostly move within a given city or commuting zone; thus, this does not bias our results because our treatment is defined at the location level. Interestingly, we find that new transport infrastructure does affect within-location firm relocation.

1.4.7 Discussion

Our findings have shown that new transport infrastructure has a positive effect on local labor productivity, and that this increase is associated to an increase in firm TFP. The positive effects of infrastructure on firm-level productivity we find are in line with those of previous findings (Holl, 2016; Holl and Mariotti, 2018; Gibbons et al., 2018). Labor and firm productivity may increase for many reasons. For example, average firm TFP can stem from better firm selection or from agglomeration externalities (Combes et al., 2012). Wan and Zhang (2017); Lee (2021); Xu and Feng (2022) provide empirical evidence that new highways incentives firm agglomeration, and Ahlfeldt and Feddersen (2018) find that infrastructure is a driver of better firm selection. Though our empirical study cannot disentangle selection from agglomeration effects, our evidence on firm entry and exit suggest that firm selection plays an important role in the overall increase in productivity.

The findings also show that new infrastructure affects firm dynamics – that is, it impacts the processes of firm entry, exit, and growth. The literature has found mixed evidence regarding infrastructure's impacts on the entry process. Audretsch et al. (2017); Gibbons et al. (2019) find that the number of firms in places with better access to infrastructure increases, mostly driven by entry; however, Chang and Zheng (2022) find no effects on entry and a decline in the number of firms in locations exposed to new transport infrastructure. In general, we do not find statistically significant effects on entry in the short run. However, we find positive effects one period after the treatment (five years later). Research documenting the effects of transport infrastructure on firm exit is scarce. We find negative effects of new highways on firm exit, which is consistent with a story in which better highways decrease trade costs and increase firm profitability and chances of survival.[12]

In the following section, we propose a model that rationalizes why new transportation infrastructure distorts firm dynamics, and how this mechanism determines location-level productivity. In the model, local labor productivity is directly determined by the composition for firms; thus, firm selection is an important channel that drives the effects of better infrastructure on economic outcomes. Although we do not model agglomeration forces directly, the model is flexible enough to account for them at no computational cost (as described in Appendix 1.C).[13]

1.5 Model

In Section 1.4, we documented that new highways increased firm-level TFP, firm entry, and the likelihood of a firm's survival. These positive effects translate into higher local labor productivity. In this section, we outline a theoretical framework that allows us to interpret these results and study the implications for aggregate output, welfare, and the spatial distribution of economic activity. To do this, we build upon an economic geography model à la Allen and Arkolakis (2014) to incorporate firm dynamics by which we mean the endogenous processes of entry, exit, and growth of heterogeneous firms in the tradition of Melitz (2003).

1.5.1 Geography

Time is discrete and indexed by t. In each period, there exists a fixed set of locations in the country denoted by $\mathcal{J} = 1, 2, \ldots, J$.[14] Locations in this economy are understood as local labor markets such as cities or commuting zones. They are interconnected by a network of highways that can be improved by building new routes or upgrading existing ones. Improvements in the highways network can reduce the minimum travel times between any two locations.[15] We

[12]See Grover Goswami et al. (2024) for the effects on competition. See Baum-Snow et al. (2024) for firm to location productivity pass-through.

[13]The main challenge is identification. It is seldom straightforward to disentangle the parameter governing agglomeration externalities from baseline productivity.

[14]For endogenous city formation see Gaubert (2018).

[15]We rule out the Braess's paradox stating that adding one or more roads to a road network may slow down overall traffic flow through it. A paper featuring congestion is Allen and Arkolakis (2022).

denote the matrix of minimum travel times between locations i and j as $\{T_{i,j,t}\}_{i,j\in\mathcal{J}}$.

We assume that the matrix of minimum travel times is sufficient to determine a geography of bilateral trade costs, denoted by $\{\tau_{i,j,t}\}_{i,j\in\mathcal{J}}$.[16] For now, we remain agnostic about the exact function that maps travel times to trade costs. From now on, for all variables with subscript (i,j), (i) denotes the origin, and (j) denotes the destination.

1.5.2 Households

At time t, the country is inhabited by an exogenous number of perfectly mobile households, denoted as $\bar{L}_t$. Households decide where to reside and how much to consume. Each household is endowed with one unit of labor, which is inelastically supplied to the local labor market at a wage rate of $w_{i,t}$. The household consumes a basket of varieties $c_{j,i,t}(n)$ produced by firm n in location j. These varieties form a composite good $C_{i,t}$ aggregated à la Dixit and Stiglitz (1977):

$$C_{i,t} = \left[\sum_{j\in J}\sum_{n\in M_{j,t}} c_{j,i,t}(n)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{1.2}$$

Where $\sigma > 1$ is the elasticity of substitution across all varieties. The price of a variety $c_{j,i,t}(n)$ is denoted by $p_{j,i,t}(n)$. Utility is derived from this basket of goods and local amenities $u_{i,t}$ according to the function:

$$U_{i,t} \equiv C_{i,t} \cdot u_{i,t} \tag{1.3}$$

Amenities rationalize why households move to certain places despite receiving lower wages. These considerations include good weather, cultural attractions, family ties, and birthplace preferences (Zerecero, 2021). The consumption basket $\{c_{j,i,t}(n)\}$ maximizes (1.3) subject to the budget constraint:

$$\sum_{j\in J}\sum_{n\in M_{j_t}} p_{j,i,t}(n)c_{j,i,t}(n) = w_{i,t} + d_{i,t} \tag{1.4}$$

Where $d_{i,t}$ denotes the dividends paid by the firms to households. We assume that all profits are collected by a central fund and then redistributed. Using the approach undertaken by Chaney (2008), each household owns $w_{i,t}$ shares of the fund, thus, income is proportional to the local wage and does not affect household's location choices. For the sake of simplicity in notation, we omit dividends from the equations.[17]

From the household's utility-maximization problem we can show that the instantaneous, indirect utility depends on the real wage $\frac{w_{i,t}}{P_{i,t}}$ and local amenities $u_{i,t}$ as follows:

$$U_{i,t} = \frac{w_{i,t}}{P_{i,t}} \cdot u_{i,t} \tag{1.5}$$

[16] This assumption is reasonable in the absence of internal tariffs.

[17] Under this assumption, we can show that the actual income is $\frac{\sigma}{\sigma-1}w_{i,t}$ which proportionally shifts welfare $U_{i,t}$ for all i.

Where $P_{i,t}$ is the standard price index of location i, defined as:

$$P_{i,t} \equiv \left[\sum_{j \in \mathcal{J}} \sum_{n \in M_{j,t}} p_{j,i,t}(n)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \tag{1.6}$$

Aggregate Marshallian demand of the $L_{i,t}$ households in location i for a variety n produced at location j is:

$$c_{j,i,t}(n) = p_{j,i,t}(n)^{-\sigma} w_{i,t} L_{i,t} P_{i,t}^{\sigma-1} \tag{1.7}$$

According to this demand function, the price elasticity is σ and $w_{i,t} L_{i,t} P_{i,t}^{\sigma-1}$ is a local demand shifter that proportionally raises demand for all local varieties. The demand function implies that households demand a positive amount of all varieties as long as there exists a firm willing to produce them.[18]

Households' location choice. Households are freely mobile and decide where to live at the beginning of every period. The value of living at location i at time t is:

$$W_{i,t} = U_{i,t} + \beta \mathbb{E}_{\Omega} \left[W_{t+1} | \Omega_t \right] \tag{1.8}$$

Where Ω_t is the aggregate state of the economy at time t, which includes all information about the distribution of prices and quantities across locations. Households discount the future at rate $\beta \in (0,1)$ forming beliefs $\mathbb{E}_{\Omega}$ through expectations that may depart from rational. The continuation value is W_{t+1}, defined as:

$$W_{t+1} = \max_{j \in \mathcal{J}} \{ W_{j,t+1} \} \tag{1.9}$$

The absence of a moving cost in the continuation value reflects the fact that households can freely move from location i to j. The location choice is then:

$$i = arg \max_{j \in \mathcal{J}} W_{j,t} \tag{1.10}$$

1.5.3 Firms

Technology. In period t, there are $M_{i,t}$ heterogeneous, risk-neutral firms at location i. They use labor to produce a single variety, indexed by $n \in M_{i,t}$, with the following constant returns to scale technology:

$$y_{i,t}(n) = \psi_{i,t}(n) \cdot l_{i,t}(n) \tag{1.11}$$

Where firm-level productivity, $\psi_{i,t}(n)$, is separable in two parts as:

$$\psi_{i,t}(n) = z_{i,t} \cdot s_i(n) \tag{1.12}$$

[18] It is straightforward to extend this framework to many sectors as in Asturias et al. (2019). This will imply having different elasticities within and across sectors. Arkolakis et al. (2019) show that heterogeneous markups may imply smaller welfare gains from trade.

Where $z_{i,t}$ is a random location-specific productivity shifter.[19] It rationalizes why the same firm would exhibit different labor productivity when situated in a different location or when experiencing distinct time periods. On the other hand, $s_i(n)$ is the idiosyncratic productivity of a firm, which is time invariant and drawn before entry.

Profit maximization. Firms operate in a monopolistic competition market and sell their products to all locations. When firm n in location i serves market j, it chooses labor, output, and prices to solve:

$$\begin{aligned} \max_{p_{i,j,t}(n),y_{i,j,t}(n),l_{i,j,t}(n)} \pi_{i,j,t}(n) = p_{i,j,t}(n)y_{i,j,t}(n) - w_{i,t}l_{i,j,t}(n) \qquad \forall j \in \mathcal{J} \\ \text{subject to} \qquad (1.7) \end{aligned} \tag{1.13}$$

Consumers at j pay $p_{i,j,t}(n) = \tau_{i,j,t} \cdot p_{i,t}(n)$, where $p_{i,t}(n)$ is the price at the location of origin. At the optimum, firms will price a constant markup over the marginal cost. That is:

$$p_{i,j,t}(n) = \left(\frac{\sigma}{\sigma - 1}\right) \frac{\tau_{i,j,t}w_{i,t}}{\psi_{i,t}(n)} \tag{1.14}$$

Optimal labor and quantities follow from (1.14) and the demand and production functions. Equation (1.14) implies that all differences in prices of the variety n are fully explained by differences in trade costs; so any reductions in trade costs will be fully passed on to consumers in the form of a lower price.

Definition 1. *Location's i **market access** is:*

$$ma_{i,t} \equiv \tilde{\sigma} \sum_{j \in \mathcal{J}} \tau_{i,j,t}^{1-\sigma} w_{j,t} L_{j,t} P_{j,t}^{\sigma-1} \tag{1.15}$$

where $\tilde{\sigma} \equiv \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}$.

Proposition 1. *Firm's (total) optimal labor demand and profits are:*

$$l_{i,t}(n) = \psi_{i,t}(n)^{\sigma-1} \cdot w_{i,t}^{-\sigma} \cdot ma_{i,t} \tag{1.16}$$

and

$$\pi_{i,t}(n) = \frac{1}{\sigma} \cdot \psi_{i,t}(n)^{\sigma-1} \cdot w_{i,t}^{1-\sigma} \cdot ma_{i,t} \tag{1.17}$$

Proof. See Appendix 1.C. □

According to (1.16) and (1.17) static optimal decisions of firms depend only on their productivity, local wages, and market access.[20]

Incumbent's problem. Firms decide to stay or exit after production takes place and profits are realized. The value of an incumbent firm in location i producing variety n is:

$$V_{i,t}(n)^I = \pi_{i,t}(n) + \beta \mathbb{E}_\Omega \left[V_{i,t+1}(n) | \Omega_t\right] \tag{1.18}$$

[19] This shifter can be further decomposed by making assumptions on, for instance, agglomeration externalities (Combes et al., 2012).

[20] Firms have no local labor market power; thus they take wages and market access as given. See Azkarate-Askasua and Zerecero (2022) for local labor market power.

Where the continuation value, normalizing the outside option for entrepreneurs to zero for all n, is:

$$V_{i,t+1}(n) = \max\{V_{i,t+1}(n)^I - f_{i,t}(n), 0\} \tag{1.19}$$

Here, $f_{i,t}(n)$ is a random operating cost drawn at the end of period t. This operating cost provides the rationale that can explain why there is not a hard productivity cut-off for exiting firms in the data; some productive firms exit, and some unproductive firms stay in the market.

Entrants' problem. At the end of period t, exogenous $M_{i,t}^{PE}$ potential entrants draw idiosyncratic productivity shocks $\{s_i(n)\}_{n\in M_{i,t}^{PE}}$, then, determine the value of entering and starting operations in $t+1$:

$$V_{i,t}(n)^E = \beta \mathbb{E}_\Omega \left[V_{i,t+1}(n)^I | \Omega_t\right] \tag{1.20}$$

Where $e_{i,t}(n)$ is a random entry cost observed before making the entry decision. Normalizing the outside option to zero, the potential entrant decides to enter in $t+1$ if:

$$V_{i,t}(n)^E - e_{i,t}(n) > 0 \tag{1.21}$$

The entry shock $e_{i,t}(n)$ explains why certain unproductive firms might enter the market while some highly productive ones might not. As the productivity draw increases, so does the value of entering the market, making it more likely for a firm to choose to enter.

1.5.4 Local labor productivity

In standard economic-geography models, production in a location takes place in a single representative firm with a production function of the form $Y_{i,t} = A_{i,t}L_{i,t}$, where $A_{i,t}$ is local labor productivity and is exogenously given and, therefore, policy invariant. The key innovation of our framework is that we allow $A_{i,t}$ to depend on local productivity shocks and the endogenous and dynamic firm composition.

Definition 2. *The endogenous* ***location-level labor productivity*** *is:*

$$A_{i,t} \equiv \left[\sum_{n\in M_{i,t}} \varphi_{i,t}(n)^{\sigma-1}\right]^{\frac{1}{\sigma-1}} \tag{1.22}$$

Notice that definition 2 is isomorphic to a framework in which a location-specific variety is produced using intermediate inputs from local firms and aggregated according to (1.2). Combining this definition with the firm's production function, the productivity of a location can be rewritten as:

$$A_{i,t} = z_{i,n} \cdot \left[\sum_{n\in M_{i,t}} s_i(n)^{\sigma-1}\right]^{\frac{1}{\sigma-1}} \tag{1.23}$$

Equation (1.23) shows that the labor productivity of a location depends on the exogenous productivity shifter $z_{i,t}$, the endogenous number of firms $M_{i,t}$, and importantly, the endogenous idiosyncratic productivity distribution $\{s_i(n)\}_{n\in M_{i,t}}$. The distribution of $s_i(n)$ is determined by the incumbent and potential entrant problems described above and evolves according to the following process:

$$\{s_i(n)\}_{n\in M_{i,t}} = \{s_i(n)\}^I_{n\in M_{i,t-1}} \cup \{s_i(n)\}^E_{n\in M_{i,t}} \tag{1.24}$$

Intuitively, the current set of producing firms is the union of the sets of surviving firms from the previous period and the potential entrants that decided to start production in t.

Proposition 2. *Output at the location level, given by $Y_{i,t} = A_{i,t}L_{i,t}$, can be decomposed as:*

$$log(Y_{i,t}) = \underbrace{log(z_{i,t})}_{\text{Technology shock}} + \underbrace{log(\tilde{s}_{i,t})}_{\text{Firm selection}} + \left(\frac{1}{\sigma-1}\right)\underbrace{log(M_{i,t})}_{\text{Varieties}} + \underbrace{log(L_{i,t})}_{\text{Total labor}} \tag{1.25}$$

Where $\tilde{s}_{i,t} \equiv \left[\frac{1}{M_{i,t}}\sum_{n\in M_{i,t}} s_i(n)^{\sigma-1}\right]^{\frac{1}{\sigma-1}}$ is the generalized mean idiosyncratic productivity of location i. All terms in the decomposition are positively valued.

Proof. Combine (1.2), (1.11), (1.15) and (1.16). □

Equation (1.25) shows that a location will produce more composite output per worker if it faces favorable exogenous technology shocks; if firm selection improves; or if many firms agglomerate in the location.

1.5.5 Equilibrium

Timing. Figure 1.5 illustrates the timing of our model. At the beginning of period t, all agents observe the realization of local amenities, productivity shocks, trade costs, and total population. As the composition of firms in period t was decided in $t-1$, local labor productivities, $A_{i,t}$, are immediately determined. Then, households determine labor supply by deciding where to live, taking prices and wages as given.

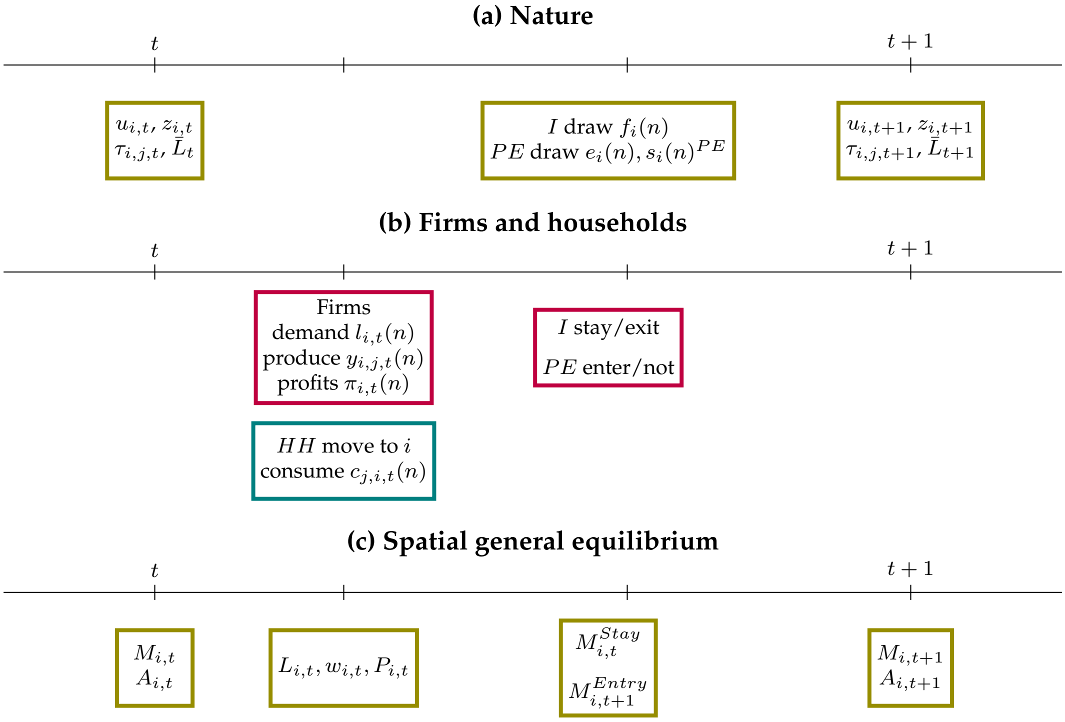
Simultaneously, firms decide their labor demand and production levels, taking market access and wages as given. Finally, profits are realized and redistributed to households. Before the end of the period, incumbent firms decide whether they will continue or exit, and potential entrants decide whether to enter or not. Once these decisions are made, the number and composition of active firms in $t+1$ are determined.

Similar to Caliendo et al. (2019), we establish a distinction between a *temporary* and a *sequential* competitive equilibrium. The *temporary* equilibrium is the solution to the multi-location internal trade model. The *sequential* equilibrium is characterized by the migration decisions of households and the entry and exit decisions of firms.

Definition 3. *Given $\bar{L}_t, u_{i,t}, z_{i,t}, \tau_{i,j,t}$, a* ***temporary equilibrium*** *are quantities $L_{i,t}, y_{i,j,t}$ and prices , $w_{i,t}, p_{i,t}(n), P_{i,t}$ such that:*

1. *Households maximize utility given by (1.4)*
2. *Firms maximize profits given by (1.13)*

Figure 1.5: Timing of the model



Notes: I=incumbents. PE=potential entrants. HH=households

3. *Wages $w_{i,t}$ clear local labor markets $\forall i \in \mathcal{J}$:*

$$L_{i,t} = \sum_{n \in M_{i,t}} l_{i,t}(n)$$

4. *Prices $p_{i,j,t}(n)$ clear good markets $\forall n \in M_{i,t}$ and $\forall i, j \in \mathcal{J}$:*

$$c_{i,j,t}(n) = y_{i,j,t}(n)$$

$$w_{i,t} L_{i,t} = \sum_{j \in \mathcal{J}} \sum_{n \in M_j} p_{i,j,t}(n) y_{i,j,t}(n)$$

Definition 4. *Given $\bar{L}_t, u_{i,t}, z_{i,t}, \tau_{i,j,t}$ and $f_{i,t}(n), e_{i,t}(n)$, a* ***sequential equilibrium*** *are quantities $L_{i,t}$, $M_{i,t}$ such that:*

1. *Migration decisions solve (1.10) and utility is equalized across locations $U_{i,t} = U_t \; \forall i \in \mathcal{J}$, moreover:*

$$\sum_{i \in \mathcal{J}} \sum_{j \in \mathcal{J}} (L_{i,t} - L_{j,t-1}) = \bar{L}_t - \bar{L}_{t-1}$$

2. *Entry and exit decisions solve (1.19) and (1.21) and:*

$$M_{i,t} = M_{i,t}^S + M_{i,t-1}^E \forall i \in \mathcal{J}$$

Where $M_{i,t}^S$ denotes the mass of surviving firms from $t-1$ to t.

Proposition 3. *The static equilibrium exists, and it is unique; therefore, the sequence of temporary equilibria exists, and is unique. Moreover, for arbitrary constants U_t and ϕ_t, the following system of equations determines the static spatial equilibrium.*

$$L_{i,t}w_{i,t}^{\sigma} = \tilde{\sigma}U_t^{1-\sigma}\sum_{j\in\mathcal{J}}\tau_{i,j}^{1-\sigma}A_{i,t}^{\sigma-1}u_{j,t}^{\sigma-1}L_{j,t}w_{j,t}^{\sigma} \tag{1.26}$$

$$w_{i,t}^{1-\sigma} = \tilde{\sigma}U_t^{1-\sigma}\sum_{j\in\mathcal{J}}\tau_{i,j}^{1-\sigma}A_{j,t}^{\sigma-1}u_{i,t}^{\sigma-1}w_{j,t}^{1-\sigma} \tag{1.27}$$

$$L_{i,t}w_{i,t}^{\sigma}A_{i,t}^{1-\sigma} = \phi_t w_{i,t}^{1-\sigma}u_{i,t}^{1-\sigma} = ma_{i,t} \tag{1.28}$$

Proof. From market clearing, the indirect utility function, and the price index, we obtain (1.26). From the price index and the indirect utility function we get (1.27). From theorems 1 and 2 in Allen and Arkolakis (2014) we know that, given $\bar{L}_t, u_{i,t}, A_{i,t}, \tau_{i,j,t}$, the sequence of static equilibrium exists, it is unique, and it satisfies 1.28. □

Proposition 4. *There is a unique allocation of workers across firms within location given by:*

$$\frac{l_{i,t}(n)}{L_{i,t}} = \left(\frac{s_i(n)}{\bar{s}_{i,t}}\right)^{\sigma-1} \tag{1.29}$$

Where $\bar{s}_{i,t} \equiv \left(\sum_{n\in M_{i,t}} s_{i,t}(n)^{\sigma-1}\right)^{\frac{1}{\sigma-1}}$.

Proof. Combine equations (1.16) and (1.28). □

According to (1.29), there is a convex relationship between a firm's relative productivity and its relative size. If firms with relatively high levels of productivity enter location i such that firm n is 1% relatively less productive, it will lose $(\sigma-1)$% of its share in the local labor force.

1.6 Calibration

In this section, we take our model to the Mexican data by using a combination of parameterization, model inversion, and internal calibration. Then, we conduct two validation exercises to test the model's predictive performance.

1.6.1 Parameterization

Time period and locations. In our model, each period spans five years, aligning with the frequency of our five census waves: 1998, 2003, 2008, 2013, and 2018. We set the 5-year discount rate to $\beta = 0.82$, consistent with an annual discount rate of 0.96. We restrict the number of locations to $\mathcal{J} = 2,463$. The rest have been excluded because they do not consistently appear in all census waves, or they have fewer than 10 firms, which addresses confidentiality concerns. Our 2,463 locations encompass 93% of all firms in 1998 and 85% in 2018.

Elasticity of substitution. We set $\sigma = 9$ for all periods, following Eaton and Kortum (2002) and Allen and Arkolakis (2014). This choice allows us to ensure comparability of our results with standard internal trade models.[21] This value is higher than what is often found in the literature (e.g Hsieh and Klenow (2009)). Lower values of σ would imply lower substitutability across goods and therefore larger gains from the reduction in trade costs.

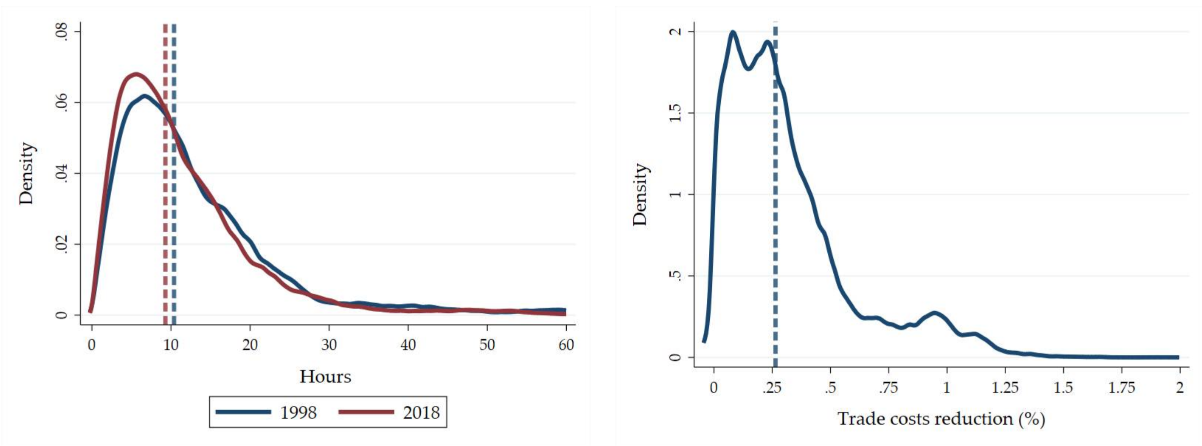
Trade costs. We estimate trade costs $\tau_{i,j,t}$ for all census waves in two-steps. First, we compute the minimum travel time between any two locations, $T_{i,j,t}$, by using the Dijkstra algorithm (Dijkstra, 1959). This algorithm discretizes the space in cells characterized by the speed of their highways. If a cell is not intersected by any highway, we assume a transit speed of 5 km/h. If it is crossed by one or more highways, the transit speed is determined by the one with the highest maximum speed, which ranges from 50 km/h to 120 km/h. We set the speed in cells forming urban agglomerations to be 30 km/h.

Once we have the minimum travel times for all pairs i, j and for all t, we compute the trade costs as in Hanson (2005) and Pérez and Sandoval (2017) assuming the following parametric form:

$$\tau_{i,j} = \begin{cases} e^{\lambda_0 + \lambda_1 T_{i,j}} & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \tag{1.30}$$

Where λ_0 represents the fixed cost of the goods leaving the location of origin, and λ_1 denotes the additional cost incurred for each additional hour of transportation time. We parameterize this function following Pérez and Sandoval (2017). They estimate $\lambda_0 = 0.0557$ and $\lambda_1 = 0.0024$ for Mexico using price data for avocados, which are a good primarily produced in a single location and sold at prices that increase with travel time. Their estimates imply that when goods leave the location of origin, prices increase immediately by 5.57% and then increase by 5.76% for every 24 hours in transit.

Figure 1.6: Travel times and reduction in trade costs, 1998-2018



(a) Hours origin-destination

(b) Trade costs reduction, 1998-2018

Notes: Figures show all origin-destination i, j combinations $(3, 234^2)$.

Figure 1.6 panel (a) shows the distribution of travel time hours for all pairs of origins and destinations in the data. In 1998, the median origin-destination travel time was 13.4 hours. This decreased to 11.6 hours in 2018. Panel (b) shows how the overall reduction in travel times

[21] In Gaubert (2018) this is calibrated to match the average revenue-to-cost margin in each sector.

affected the implied trade costs. The median origin destination pair (i, j) saw a reduction of 0.26% in trade costs.

1.6.2 Labor and wage paths

Labor and wages. We assume that the observed geographical distributions of wages $w_{i,t}$ and labor $L_{i,t}$ across locations are equilibrium outcomes of the model. We measure the local labor distribution $\{L_{i,t}\}_{i\in\mathcal{J}}$ as the number of workers reported in the census.

The distribution of local wages $\{w_{i,t}\}_{i\in\mathcal{J}}$, is obtained by residualizing wages in two steps. First, we compute local average wages $\bar{w}_{i,t}$ as total wage bill over the number of workers. And second, we regress it on local observable characteristics that are not accounted for by our model and use the estimated residuals as local wages. The regression model is:

$$\bar{w}_{i,t} = \beta_0 + \beta_1\% \text{ educ}_{i,t} + \beta_2\% \text{ manuf}_{i,t} + \beta_3 \text{K/L}_{i,t} + \beta_4\% \text{ inf}_{i,t} + \epsilon_{i,t} \tag{1.31}$$

Where i denotes the location and t the census year. The regression accounts for heterogeneity in education, industrial composition, capital intensity, and informality. Figure 1.18 in Appendix 1.C shows the correlation between residualized wages and local population. Wages in large locations are higher even after controlling by observable characteristics. This is in line with a story in which locations with highly productive firms increase both local labor productivity and wages, and thus attract more workers.

1.6.3 Model inversion

Local amenities and productivity. We invert the model to retrieve the distribution of local amenities $u_{i,t}$ and local labor productivity $A_{i,t}$. For a given geography of trade costs, differences in amenities are identified from differences in population in locations with similar wages. On the other hand, differences in labor productivity are identified from differences in labor income in locations with similar amenities. Formally, (1.32) and (1.33) retrieve amenities and productivities from an observed distribution of trade costs, local labor, and wages.[22]

$$u_{i,t}^{1-\sigma} = \frac{\tilde{\sigma} U_t^{1-\sigma}}{\phi_t} \sum_{j\in\mathcal{J}} \tau_{i,j,t}^{1-\sigma} w_{i,t}^{\sigma-1} w_{j,t}^{\sigma} L_{j,t} u_{j,t}^{\sigma-1} \qquad \forall i \in \mathcal{J} \tag{1.32}$$

$$A_{i,t} = \left[\frac{1}{\phi_t} L_{i,t} w_{i,t}^{2\sigma-1} u_{i,t}^{\sigma-1}\right]^{\frac{1}{\sigma-1}} \qquad \forall i \in \mathcal{J} \tag{1.33}$$

To determine ϕ_t we use the equilibrium in the labor market $\bar{L}_t = \sum_{i\in\mathcal{J}} L_{i,t}$:

$$\phi_t = \bar{L}_t \left(\sum_{i\in\mathcal{J}} w_{i,t}^{1-2\sigma} u_{i,t}^{1-\sigma} A_{i,t}^{\sigma-1}\right)^{-1} \tag{1.34}$$

Figure 1.7 illustrates the correlation between amenities, labor productivity, population, and wages. Two contrasting cases, Tijuana and Merida, highlight how these variables interact.

[22]Endogenous constants U_t and ϕ_t are not identified in levels. We normalize them to one at baseline.

Tijuana is a dangerous city located in the northern Mexican desert, while Merida, situated near the Caribbean Sea, is renowned for its safety. Tijuana has limited local amenities in comparison to Merida; nonetheless, Tijuana has higher population and wage levels than Merida. This is explained by Tijuana's higher local labor productivity, which is driven by its highly productive firms in the export-oriented manufacturing sector.

Figure 1.7: Amenities, productivity, and equilibrium outcomes, 2018

(a) Productivity and employment

(b) Productivity and wages

(c) Amenities and employment

(d) Amenities and wages

Notes: Marker size denotes the number of firms in the location.

1.6.4 Internal calibration

Once we have fully characterized the path of aggregate location-level equilibrium outcomes, we exploit the microdata to determine the primitives that govern firm dynamics in the model. These are the path of location-level productivity shocks, the initial distribution of idiosyncratic productivities, the entry- and exit-cost distributions, and the path of potential entrants.

Location-specific productivity shock. From (1.23) and defining

$$\bar{s}_{i,t} \equiv \left[\sum_{n \in M_{i,t}} s_i(n)^{\sigma-1} \right]^{\frac{1}{\sigma-1}}, \tag{1.35}$$

we solve for the location-specific productivity shock as follows:

$$z_{i,t} = \frac{A_{i,t}}{\bar{s}_{i,t}} \tag{1.36}$$

Computation of $z_{i,t}$ requires first $A_{i,t}$ which comes from model inversion described above; and second, the distribution of firm-level idiosyncratic productivity, which is a sequential equilibrium outcome. This distribution $\{s_{i,t}(n)\}_{n \in M_{i,t}}$ $\forall i \in \mathcal{J}$ depends on the initial distribution of idiosyncratic productivities and the entry and exit costs' stochastic distributions.

Initial idiosyncratic productivity distribution. To identify the initial distribution of $s_{i,t}(n)$ we assume that the economy reached the steady state in 1998 starting from an arbitrary point in the past. Then, we exploit the result that in equilibrium:

$$\frac{l_{i,t}(n)}{L_{i,t}} = \left(\frac{s_{i,t}(n)}{\bar{s}_{i,t}}\right)^{\sigma-1} \tag{1.37}$$

Thus, the observed distribution of firm-level labor demand is fully informative about the initial idiosyncratic productivity distribution. More precisely, if $\tilde{l}_{i,t}(n) = \left(\frac{1}{\sigma-1}\right)[log(l_{i,t}(n)) - log(L_{i,t})]$ follows an arbitrary distribution $F(\mu_{\tilde{l}}, \sigma_{\tilde{l}})$, then, $s_{i,t}(n)$ follows $F(\mu_{\tilde{l}} + log(\bar{s}_{i,t}), \sigma_{\tilde{l}})$. From firm-level data we compute $\mu_{\tilde{l}}, \sigma_{\tilde{l}}$, and then we solve the fixed-point problem until $\bar{s}_{i,t}$ is consistent with the equilibrium condition.[23]

Potential entrants' productivity distribution. The distribution of idiosyncratic productivities is governed by $F(\mu_s, \sigma_s)$. Assuming that we know $F(\mu_f, \sigma_f)$ and $F(\mu_e, \sigma_e)$, we estimate the parameters μ_s, σ_s by solving the following problem:[24]

$$\{\hat{\mu}_s, \hat{\sigma}_s\} = arg \min_{\mu_s, \sigma_s} \sum_{i \in \mathcal{J}} \sum_{n \in M_i} \cdot \left[log(l_i(n)^{data}) - log(l_i(n)^{model})\right]^2 \tag{1.38}$$

Here, $l_i(n)$ is the number of workers in a firm in the data, and $l_i(n)^{model}$ is the labor demand in the model according to Equation 1.16. Intuitively, conditional on a set of values for the entry and exit costs, the optimal estimators of μ_s, σ_s are the ones that minimize the square percentage deviations in labor demands observed in the data and the ones implied in the model.

Exit costs. We estimate the exit-cost parameters as follows: first, recall that a firm at the end of period t stays in the market for period $t+1$ if the expected continuation value in $t+1$ minus a cost shock observed at the end of t is higher than the outside option, which we normalize to zero. Denote the continuation value as:

$$x_{i,t}(n) = \beta \mathbb{E}_{\Omega} \left[V_{i,t+1}(n) | \Omega_t\right] \tag{1.39}$$

Suppose that the cost shock, denoted as $f_{i,t}(n)$, comes from a Gumbel probability distribution $G(.)$. The survival probability of a firm is then:

$$\lambda(x_{i,t}(n)) = \mathbb{P}[x_{i,t}(n) > f_{i,t}(n)] = G(x_{i,t}(n)) \tag{1.40}$$

[23] In the quantitative section we assume $F(.)$ is log normal and that $\mu_{\tilde{l}}, \sigma_{\tilde{l}}$ are location specific.

[24] We need to add more details on the definition and existence of a steady state.

Denoting the location parameter μ_f and the spread parameter σ_f, we obtain:

$$\lambda(x_{i,t}(n)) = e^{-e^{-\left(\frac{x_{i,t}(n)-\mu_f}{\sigma_f}\right)}} \tag{1.41}$$

To compute $x_{i,t}(n)$ we assume that firms form myopic expectations denoted as $\tilde{\mathbb{E}}$ about the future-state space Ω_{t+1}. This implies that $\tilde{\mathbb{E}}_\Omega \left[V_{i,t+1}(n)|\Omega_t\right] = V_{i,t}(n)$. Then, the survival probability is the solution to the non-linear system given by equations 1.17 and 1.41, which gives:

$$\lambda(w_{i,t}l_{i,t}(n)) = \frac{1}{\beta} - \frac{\frac{1}{\sigma-1}w_{i,t}l_{i,t}(n)}{\mu_s - \sigma_s log[-log[\lambda(w_{i,t}l_{i,t}(n))]]} \tag{1.42}$$

Equation 1.42 shows that there is a non-linear mapping between the firm-level equilibrium wage bill $w_{i,t}l_{i,t}(n)$ and the survival probability $\lambda(w_{i,t}l_{i,t}(n))$. We leverage this relationship to retrieve the cost-shock-distribution parameters μ_f, σ_f by solving for the parameters of the cost-shock distribution that will solve the minimization problem:

$$\{\hat{\mu}_f, \hat{\sigma}_f\} = arg \min_{\mu_f, \sigma_f} \sum_{i \in \mathcal{J}} \sum_{n \in M_i} \cdot \left[\lambda(w_{i,t}l_{i,t}(n))^{data} - \lambda(w_{i,t}l_{i,t}(n))^{model}\right]^2 \tag{1.43}$$

Problem (1.43) requires that the full mapping between wage bill and exit rates are defined in the data. Since the data are granular, we approximate this relationship by grouping all wage-bill values in percentiles and then computing the associated exit rate. Finally, we approximate this relationship with a polynomial fit, using this continuous approximation as the values targeted by the minimization problem.

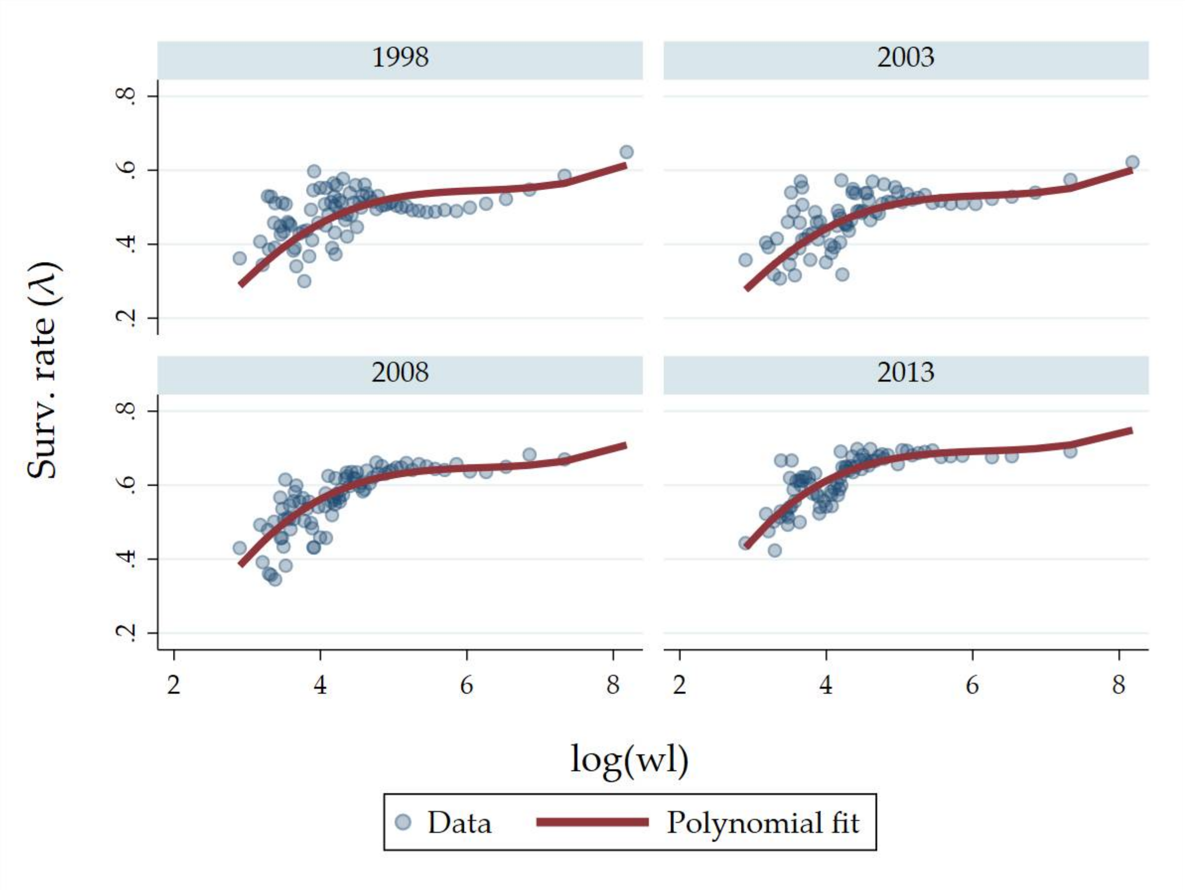
Figure 1.8 shows the polynomial fit ant the survival rates in the data. Notice that survival rates are concave for the low and middle sections of the wage-bill distribution and convex for the high end. This implies that when very large firms shrink, their survival rates decrease faster than when small firms in terms of wage bill do.

Potential entrants and entry costs. At every period we observe in the data the productivity distribution of entrants and their number. However, since by definition we do not observe the potential entrants, there are infinitely many combinations of potential entrant distributions and entry costs that rationalize the observed entrants in the data.

To address this problem we assume first that entry costs $e_i(n)$ are drawn from the same distribution as exit costs $f_i(n)$. Then, for a given productivity distribution of potential entrants, we back up the mass of potential entrants $\{M_{i,t}^{PE}\}_{i\in\mathcal{J}}$ by solving their entry problem until the implied number of entrants $\{M_{i,t}^{E}\}_{i\in\mathcal{J}}$ plus the survivors $\{M_{i,t}^{S}\}_{i\in\mathcal{J}}$ is equal to the number of firms observed in the next period $\{M_{i,t+1}\}_{i\in\mathcal{J}}$.

Finally, to recover the parameters governing the productivity distribution of potential entrants we assume that they follow a process $F(\mu_E, \sigma_E)$. Then we solve their entry problem, combine these entrants with the survivors, and verify if this productivity distribution is consistent with the one observed in the next period. We iterate on μ_E, σ_E until we reach convergence.

Figure 1.8: Wage bill (wl) and survival rate λ in the data



Notes: Each dot is a percentile p in the wage-bill distribution. The polynomial fit of degree d estimated with the ordinary- least-squares (OLS) model: $log\left(\frac{\lambda_{p,t}}{1-\lambda_{p,t}}\right) = \sum_d \gamma_d log(wL_{p,t})^d + \gamma_t + \varepsilon_{p,t}$. For $d = 3$, $\gamma_1 = 3.206$, $\gamma_2 = -0.512$, $\gamma_3 = 0.028$.

1.6.5 Model validation

Local productivity. Local labor productivity $A_{i,t}$ is identified without production data. As a validation exercise, we show that its correlation with its data counterpart, based on firm-level output data, is strong. We do this by computing $\hat{A}_{i,t}$ as in (1.22), with $\hat{\varphi}_{i,t}(n)$ estimated as value added per worker.

Figure 1.19 in the appendix shows that, for all years, the R^2 of regressing model-implied and empirical local labor productivity is close to 0.8. This suggests that the model-implied local labor productivity captures most of the variation in the data. The remaining 0.2 of the variation comes from mechanisms absent in out model, such as industrial heterogeneity or spatial frictions in human capital mobility.

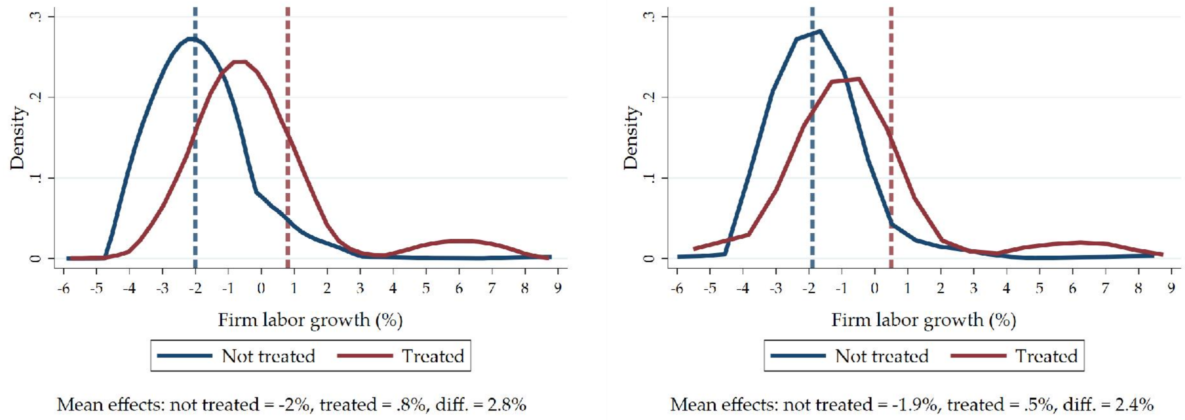
Natural experiment replication. We further validate the model by replicating the natural experiment from Section 1.4 inside the model and showing that it provides similar point estimates. We do this by creating a counterfactual scenario in which we effectively shut down all new highways, eliminating all plans that were constructed from 2013 to 2018. We then compare outcomes from the data and from this counterfactual exercise for both treated and untreated groups.[25]

[25] We do not limit the exercise to eliminating only highways from the construction plans; this is because our em-

Table 1.6 column (3) shows that the empirical point estimate is 1.6% in a 90% confidence interval of [0.1%, 3.1%]. Figure 1.9 panel (a) shows that the associated effect in the model is 2.8%, which falls within the confidence interval. We interpret this as reasonable evidence that the model is capable of capturing the observed behavior in the data.

Furthermore, we use the model to argue that the *"no anticipation of the treatment"* assumption in the empirical exercise implies an underestimation of our estimates. To address this issue, we allow firms in 2013 to expect that all announced construction plans will be built and to make their surviving decisions accordingly. We then compare treated and untreated groups in 2018. Panel (b) reveals that if we allow firms to react to the announcement, the net effect is 0.4 percentage points smaller, which is still within the confidence interval but closer to the empirical estimate. This result suggests that our empirical estimates are likely to be a lower bound of the true effect.

Figure 1.9: Effects of new highways 2013-2018 model vs natural experiment



(a) Firms do not expect plans to be executed

(b) Firms expect plans to be fullyl executed

Notes: The figure shows the replication of natural experiment for 2013-2018 construction plans. In panel (a) incumbent firms do not expect plans to be executed. In panel (b) incumbent firms expect all plans to be executed. Vertical dashed lines denote the corresponding average effect.

1.7 Quantitative results

1.7.1 Contribution of highways to welfare and growth

Between 1998 and 2018, the network of paved roads and highways in Mexico expanded from approximately 100,000 kilometers to nearly 200,000 kilometers. In this section, we show that this expansion produced modest welfare and income gains but high reallocation of economic activity across locations. We then show that firm dynamics were a key driver of both the aggregate and distributional effects. To analyze these dynamics we construct a counterfactual scenario in which the trade geography remains at 1998 levels, and we then recalculate the growth trajectory using our model. We interpret the difference between this counterfactual scenario and the path in the data as the contribution of highways that were constructed between 1998 and 2018.

pirical estimates may also capture effects from secondary roads or other highways that influence both the treatment and control groups.

Welfare and aggregate effects. New highways built from 1998 to 2018 increased welfare, real income, and productivity. Table 1.7 shows the comparisons of results from the counterfactual scenario in which the highway network remained as it was in 1998 and those outcomes as shown in our data that actually occurred in 2018. Welfare is 0.44% higher. Real income is 0.64% higher. And aggregate productivity is 0.13% higher. The number of firms, however, is 0.10% lower.

Welfare gains are entirely explained by the increase in real income as amenities are exogenous. The findings of Allen and Arkolakis (2014) serve as a reference for the extent of welfare gains; they document that the entire interstate highways network in the US increased welfare by 1.3%. As our model reveals, real income rises for two reasons. First, labor productivity improves, leading to higher nominal wages. Second, reductions in trade costs drive down prices of goods, as indicated by reductions in local price indices. The increase in productivity is explained by positive firm selection, driven by higher survival and entry rates of productive firms. Then, too, a more efficient transportation system requires fewer firms in the aggregate, as lower trade costs allow fewer firms to serve more markets.

Table 1.7: Gains from highways

Year	(1) Welfare	(2) Real income	(3) Productivity	(4) Firms
1998	0.00%	0.00%	0.00%	0.00%
2003	0.13%	0.09%	0.04%	-0.04%
2008	0.24%	0.36%	0.25%	-0.02%
2013	0.40%	0.40%	0.22%	-0.07%
2018	0.44%	0.64%	0.13%	-0.10%

Notes: Gains measure how much higher outcomes are with respect to a counterfactual in which none of the new highways after 1998 were built. L denotes total labor productivity as in (1.22). wL/P is total real remunerations. M is the total number of firms.

Distributional effects. Aggregate results hide important distributional effects across space. To illustrate this point, Table 1.8 shows gains at the 25th, 50th and 75th percentile levels in labor, real income, labor productivity, and number of firms that result from new infrastructure across all locations from 1998 to 2018.

People migrate until utility is equalized; thus, there is no dispersion in welfare gains. This leads to net migration-implied population gains of at least 5% above the 75th percentile and similar levels of losses below the 25th percentile. Due to these population losses, as shown by Column (2), real income fell in more than half of the locations, while, at the same time, real income for those whose earnings were in the top 75th percentile rose by at least 5%.

Firms react differently across space to new transport infrastructure. As shown in Column (3), labor productivity decreased in half of the locations due to exit of productive firms. Even though, in the aggregate, better transport infrastructure implies that fewer firms are needed, this is mostly driven by net exits of firms in locations locations that are farther away from the highway network. As Column (4) shows, there was a net decrease in the number of firms in more than half of locations, but a net increase of at least 2.34% in a quarter of locations.

To understand the geographical concentration of these heterogeneous gains, we show in Figure 1.10 the state-average impacts on key economic outcomes.[26] In general, the states that experi-

[26] We calculate these averages weighted by population. State-level averages are used for clarity in presentation.

Table 1.8: Distribution of impacts from new highways

Percentile	(1) Labor	(2) Real income	(3) Productivity	(4) Firms
25th	-5.76%	-5.32%	-0.60%	-2.65%
50th	-0.94%	-0.50%	0.01%	-0.12%
75th	5.17%	5.60%	0.58%	2.34%

Notes: The table shows the impacts on labor, real income, productivity and the number of firms from the new highway network by comparing outcomes with those that would have emerged in a counterfactual scenario in which the highway network had remained unchnaged since 1998. L denotes total labor productivity as in (1.22). wL/P is total real income. M is the total number of firms.

enced the largest gains are those situated near to the main port of entry to California (Tijuana) or to the Caribbean Sea, the largest tourist hub (Cancun). The states that saw moderate gains are those located near the port of entry to New Mexico and Texas (Juarez and Nuevo Laredo) or to the main sea ports connecting Mexico to Asia (Manzanillo) and Europe (Veracruz). The remaining states mostly incurred losses, indicating that more economic activity would have been concentrated there in the absence of the new transport infrastructure.

Panel (a) in Figure 1.10 shows that consumers in better connected areas can purchase goods at prices that are as much as 3% lower. According to panel (b), demand for goods produced also increases by up to 5% these locations. Panels (c) and (d) further reveal that these effects result in positive net population growth ranging from 10% to 40%, along with similar real revenue gains. In panel (e), it is shown that in these locations, the introduction of new highways increases the number of firms by 1% to 5%. While these firms may vary in productivity levels, the addition of the new highways is predominantly productivity enhancing (see panel (f)). The opposite holds true for areas with limited exposure to new transport infrastructure; these firms largely experience .

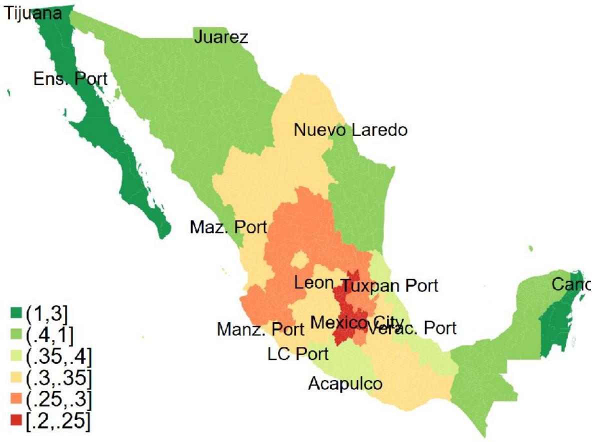
Contribution of firm dynamics. We quantify the role played by firm dynamics in the aggregate and distributional effects of new transport infrastructure in two ways. First, we compare the effects of highways on economic outcomes in our model to the effects predicted by standard trade models, which assume a static economy and exogenous local productivity. Notice that our model collapses to this framework by assuming an infinitely lived, single representative firm by location (Allen and Arkolakis, 2014).

Table 1.9 shows welfare and income gains when we abstract from firm dynamics. We omit gains in productivity and the number of firms because they are zero by definition in the absence of dynamic firm behavior. Column (1) shows that welfare gains in the absence of firm dynamics are slightly smaller. In terms of the impacts on welfare, this result suggests that the reduction in trade costs, not local productivity, matters the most for individuals. A key driver of this result is the assumption of free mobility. In terms of real income, firm dynamics play a bigger role. When we allow for firm dynamics, income gains in 2003 are 0.09%, while a standard model would imply gains of 0.05%. This means that 55% of the real revenue gains come from the reduction in trade costs induced by better highways, and that the remaining 45% of real income gains come from local productivity gains driven by firm dynamics. The contribution of firm dynamics is 11% for 2008, 16% for 2013 and 7.6% for 2018.

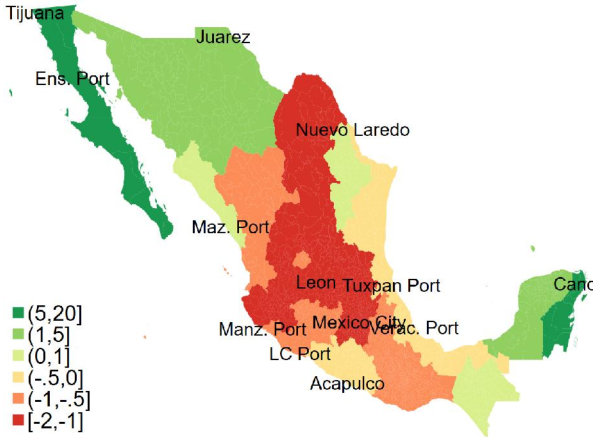
A model overlooking firm dynamics not only underestimates gains from the construction of

Figure 1.20 in the Appendix shows maps with impacts at the location level.

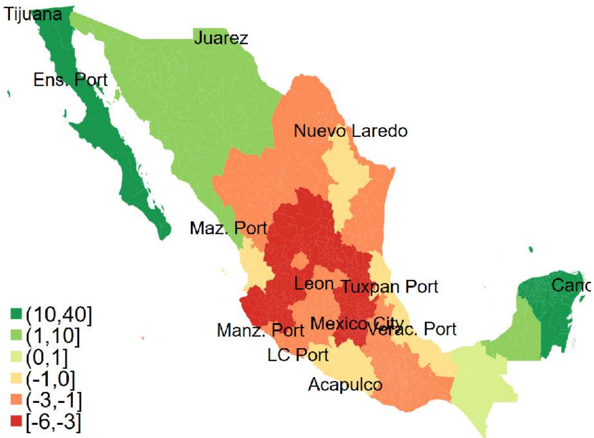
Figure 1.10: Average impacts from the highway network expansion (1998-2018)



(a) Impacts on the price index (%)

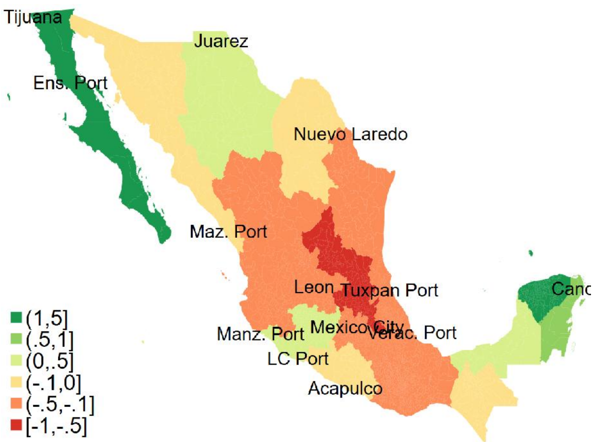


(b) Impacts on market access (%)

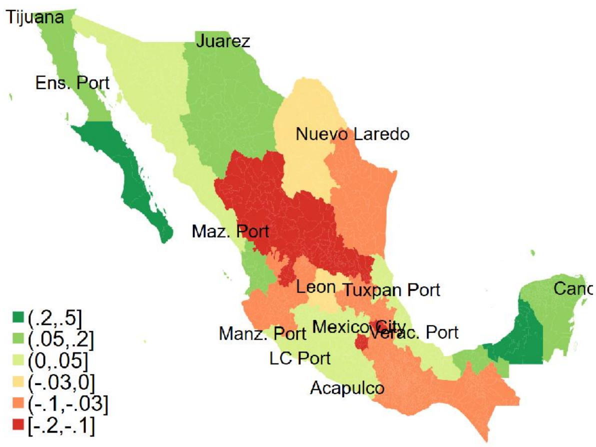


(c) Impacts on employment (%)

(d) Impacts on real revenue (%)



(e) Impacts on the number of firms (%)



(f) Impacts on productivity (%)

Notes: The maps show the impacts that stem from expanding the highways network over the period from 1998 to 2018. Changes are shown at the state level and are calculated by averaging locations weighted by population. Impacts at the municipal level are shown in Figure 1.20 in Appendix 1.C.

Table 1.9: Gains from highways, without including firm dynamics

Year	(1) Welfare	(2) Real income
1998	0.00%	0.00%
2003	0.11%	0.05%
2008	0.24%	0.32%
2013	0.39%	0.34%
2018	0.44%	0.59%

Notes: The table shows how much welfare and real income increased following the building of highways over the 1998-2018 period with respect to a counterfactual scenario that would have emerged had no new highways been built. The productivity and number of firms in a given location are kept fixed.

new highways but also their dispersion. Table 1.10 shows that the interquartile range of labor gains is 9.98%, compared to 10.93% in our baseline model. Similarly, for real income gains. This result suggests that firm dynamics are a force for spatial divergence when new transport infrastructure is unequally targeted across space.

Table 1.10: Distribution of gains from highways, without including firm dynamics

Percentile	(1) Labor	(2) Real income
25th	-5.32%	-4.87%
50th	-0.72%	-0.27%
75th	4.67%	5.12%

Notes: The table shows how much welfare and real income increased following the building of highways over the 1998-2018 period with respect to a counterfactual scenario that would have emerged had no new highways been built. The productivity and number of firms in a given location are kept fixed.

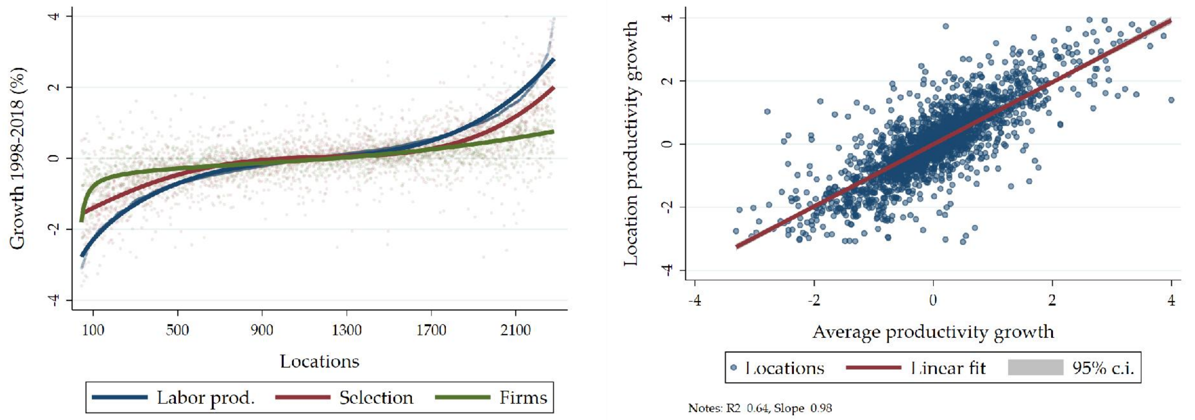
Second, we compute the extent to which the increase in local labor productivity induced by transport infrastructure can be attributed to firm selection or net firm entry. Notice that in a model without firm dynamics both are zero. Equation (1.25) implies that:

$$\begin{aligned}\Delta log(A_{i,t})^{baseline} - \Delta log(A_{i,t})^{no\ new\ highways} = \Delta log(\tilde{s}_{i,t})^{baseline} - \Delta log(\tilde{s}_{i,t})^{no\ new\ highways} \\ + \Delta log(M_{i,t})^{baseline} - \Delta log(M_{i,t})^{no\ new\ highways}\end{aligned} \tag{1.44}$$

Equation (1.44) captures the fact that some firm selection and net firm entry would have taken place in a counterfactual with no new highways over the period from 1998 to 2018. The difference between this counterfactual and what we observe in the data captures the responses of firms to new infrastructure.

Figure 1.11 shows the decomposition in (1.44) for all locations. Panel (a) shows that, in half the locations, new transportation infrastructure decreased labor productivity. Moreover, for most of the locations a larger portion of the total change is explained by firm selection rather that by net firm entry. In panel (b), we regress labor productivity growth on firm-selection growth induced by new highways over the period from 1998 to 2018. According to the R^2, 64% of the variation in labor productivity growth is explained by variation in firm selection.

Figure 1.11: Decomposition of local productivity growth induced by new highways (1998-2018



(a) Decomposition of productivity growth

(b) Correlation between location productivity and firm selection

Notes: Productivity is measured by $A_{i,t}$. Selection is measured by average idiosyncratic productivity $\tilde{s}_{i,t}$. Firms are measured by $M_{i,t}$. Dots show the J locations, and lines show a polynomial fit of degree 3.

1.7.2 A more ambitious infrastructure policy

Between 1998 and 2018 the paved roads network in Mexico doubled. The median origin-to-destination travel time (to drive from municipalities in Mexico to Mexico City) fell by 13% in 20 years, from 13.4 to 11.6 hours. Mexico has 1.4 meters of paved roads per capita. This is one-tenth of the equivalent figure for the US, Mexico's neighbor and largest trading partner. The disparity raises the question: What would have happened if infrastructure investments had been more ambitious over the period we study?

We use our calibrated model to answer this question by focusing on an alternative infrastructure policy in which the percentage reduction in travel times is twice as great as the levels calculated by using the 1998-2018 data; we then compare the likely outcomes from the two counterfactual scenarios: 1) the scenario in which the speed made possible from the highway network is twice that of the speed possible from the network that was constructed by 2018, and 2) the scenario that would have likely occurred in 2018 had no new highways been built after 1998.

Table 1.11: Likely impacts of a highway network twice as fast as the 2018 network

Year	(1) Welfare	(2) Total wL/P	(3) Total A	(4) Total M
1998	0.00%	0.00%	0.00%	0.00%
2003	0.21%	0.10%	0.01%	-0.08%
2008	0.45%	0.49%	0.28%	-0.05%
2013	0.59%	0.56%	0.24%	-0.12%
2018	0.84%	1.10%	0.14%	-0.12%

Notes: The table compares the impacts from two counterfactual scenarios, one in which journeys are twice as fast as those made possible by the network that existed in 2018, and one in which the highway network in 2018 remained the same as it had been in 1998.

Table 1.11 shows the results of this experiment. Column (1) shows that welfare and real revenue gains in 2018 would have been nearly double the level of gains from the actually built highways. Although labor productivity would be higher with a speedier highway network than the one that was constructed, the difference is small.

Table 1.12: Distribution of impacts if the highways were twice as fast as the 2018 network

Percentile	(1) Labor	(2) Real income	(3) Productivity	(4) Firms
25th	-5.14%	-5.55%	-0.62%	-2.64%
50th	0.60%	0.19%	-0.02%	0.00%
75th	5.62%	5.21%	0.59%	2.47%

Notes: The table compares impacts from two counterfactual scenarios: one in which the highway network is twice as fast as the 2018 network that was constructed, and one in which no new highways were built after 1998. L denotes total labor productivity as in (1.22). wL/P denotes total real income. M denotes the total number of firms.

Finally, Table 1.12 shows that, although the welfare and real income gains are larger, the unequal distribution of these benefits is preserved. This exercise highlights that proportionally improving the highways network has aggregate benefits but no effects on regional convergence.

1.8 Conclusion

This paper reveals that firm dynamics are a key determinant of the aggregate and distributional effects of new transportation infrastructure. We empirically document that new transportation infrastructure increases labor productivity, firms' total factor productivity, and entry and exit rates of firms.

We introduce a novel, spatial general-equilibrium model with heterogeneous firm dynamics to show that infrastructure policies affect aggregate income an welfare in two ways. The first is a direct effect: better transportation infrastructure reduces trade costs for goods, which is transmitted to consumers in the form of lower prices, and to firms as higher demand. This second is an indirect effect: new transport infrastructure increases entry and survival of productive firms in locations that are better integrated into the transportation network; this translates into higher labor productivity, income, and welfare.

These effects, however, are unequally distributed across space and among income levels. Regions close to the US border, to sea ports, and to tourist hubs are better exposed to new transportation infrastructure so they disproportionally benefit the most. Besides having greater market access, these regions also attract and keep productive firms. The opposite is true for less expose locations, mostly concentrated in the center of the country.

All in all, transport infrastructure is likely to have stronger distributional than aggregate effects, especially in situations in which the highways network is underdeveloped, as is the case in Mexico. An interesting avenue for future research could examine whether a place-based system of taxes and transfers can help mitigate the negative effects of low infrastructure investments in remote locations.

Bibliography

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Ahlfeldt, G. M. and Feddersen, A. (2018). From periphery to core: measuring agglomeration effects using high-speed rail. *Journal of Economic Geography*, 18(2):355–390.

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Allen, T. and Arkolakis, C. (2022). The welfare effects of transportation infrastructure improvements. *The Review of Economic Studies*, 89(6):2911–2957.

Arkolakis, C., Costinot, A., Donaldson, D., and Rodríguez-Clare, A. (2019). The elusive pro-competitive effects of trade. *The Review of Economic Studies*, 86(1):46–80.

Asturias, J., García-Santana, M., and Ramos, R. (2019). Competition and the welfare gains from transportation infrastructure: Evidence from the golden quadrilateral of india. *Journal of the European Economic Association*, 17(6):1881–1940.

Audretsch, D. B., Dohse, D., and dos Santos, J. P. (2017). Do toll-free highways foster firm formation and employment growth? results from a quasi-natural experiment. Technical report, Kiel Working Paper.

Azkarate-Askasua, M. and Zerecero, M. (2022). The aggregate effects of labor market concentration. *Available at SSRN 4323492*.

Baker, A. C., Larcker, D. F., and Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395.

Banerjee, A., Duflo, E., and Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics*, 145:102442.

Baum-Snow, N., Gendron-Carrier, N., and Pavan, R. (2024). Local productivity spillovers. *Nathaniel Baum-Snow Nicolas Gendron-Carrier Ronni Pavan AMERICAN ECONOMIC REVIEW. https://www. ae aweb. org/articles*.

Blankespoor, B., Bougna, T., Garduno Rivera, R., and Selod, H. (2017). Roads and the geography of economic activities in mexico. *World Bank Policy Research Working Paper*, (8226).

Borusyak, K. and Hull, P. (2020). Non-random exposure to exogenous shocks: Theory and applications. Technical report, National Bureau of Economic Research.

Busso, M., Fazio, M., and Algazi, S. (2012). (in) formal and (un) productive: The productivity costs of excessive informality in mexico.

Busso, M., Fentanes, O., and Levy Algazi, S. (2018). The longitudinal linkage of mexico's economic census 1999-2014. *Manuscript, Inter-American Development Bank, Washington, DC*.

Caliendo, L., Dvorkin, M., and Parro, F. (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. *Econometrica*, 87(3):741–835.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Chaney, T. (2008). Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review*, 98(4):1707–1721.

Chang, Z. and Zheng, L. (2022). High-speed rail and the spatial pattern of new firm births: Evidence from china. *Transportation Research Part A: Policy and Practice*, 155:373–386.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594.

Dávila, E., Kessel, G., and Levy, S. (2002). El sur también existe: un ensayo sobre el desarrollo regional de méxico. *Economía Mexicana Nueva Época, volumen XI, número 2, 2do semestre de 2002, pp 205-260*.

Dijkstra, E. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271.

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

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

Durán-Fernández, R. and Santos, G. (2014). Road infrastructure spillovers on the manufacturing sector in mexico. *Research in Transportation Economics*, 46:17–29.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Gandhi, A., Navarro, S., and Rivers, D. A. (2020). On the identification of gross output production functions. *Journal of Political Economy*, 128(8):2973–3016.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review*, 108(11):3117–53.

Gibbons, C. E., Suárez Serrato, J. C., and Urbancic, M. B. (2018). Broken or fixed effects? *Journal of Econometric Methods*, 8(1):20170002.

Gibbons, S., Lyytikäinen, T., Overman, H. G., and Sanchis-Guarner, R. (2019). New road infrastructure: the effects on firms. *Journal of Urban Economics*, 110:35–50.

Grover Goswami, A., Maloney, W., and O'Connell, S. D. (2024). Which firms drive the gains from connectivity and competition? the impact of india's golden quadrilateral across the firm life cycle.

Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration. *Journal of international economics*, 67(1):1–24.

Holl, A. (2016). Highways and productivity in manufacturing firms. *Journal of Urban Economics*, 93:131–151.

Holl, A. and Mariotti, I. (2018). Highways and firm performance in the logistics industry. *Journal of Transport Geography*, 72:139–150.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448.

Kleinman, B., Liu, E., and Redding, S. J. (2023). Dynamic spatial general equilibrium. *Econometrica*, 91(2):385–424.

Lee, J. K. (2021). Transport infrastructure investment, accessibility change and firm productivity: Evidence from the seoul region. *Journal of Transport Geography*, 96:103182.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The review of economic studies*, 70(2):317–341.

Levy, S. (2018). *Under-rewarded efforts: The elusive quest for prosperity in Mexico*. Inter-American Development Bank.

Lindenlaub, I., Oh, R., and Peters, M. (2022). Firm sorting and spatial inequality. Technical report, National Bureau of Economic Research.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *econometrica*, 71(6):1695–1725.

OECD (2020). *Transport Bridging Divides*.

Olley, S. and Pakes, A. (1992). The dynamics of productivity in the telecommunications equipment industry.

Pérez, F. and Sandoval, A. (2017). Short-run market access and the construction of better transportation infrastructure in mexico. *Economía*, 18(1):225–250.

Redding, S. J. (2016). Goods trade, factor mobility and welfare. *Journal of International Economics*, 101:148–167.

Redding, S. J. and Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. *Handbook of regional and urban economics*, 5:1339–1398.

Shiferaw, A., Söderbom, M., Siba, E., and Alemu, G. (2015). Road infrastructure and enterprise dynamics in ethiopia. *The Journal of Development Studies*, 51(11):1541–1558.

Wan, G. and Zhang, Y. (2017). The direct and indirect effects of infrastructure on firm productivity: Evidence from manufacturing in the people's republic of china. Technical report, ADBI Working Paper.

Xu, M. and Feng, Y. (2022). How transportation infrastructure affects firm productivity? evidence from china. *China Economic Quarterly International*, 2(1):55–69.

Zerecero, M. (2021). The birthplace premium. Technical report, mimeo.

Zhou, J. (2023). The impacts of highways on firm size distribution: Evidence from china. *Growth and Change*, 54(2):482–506.

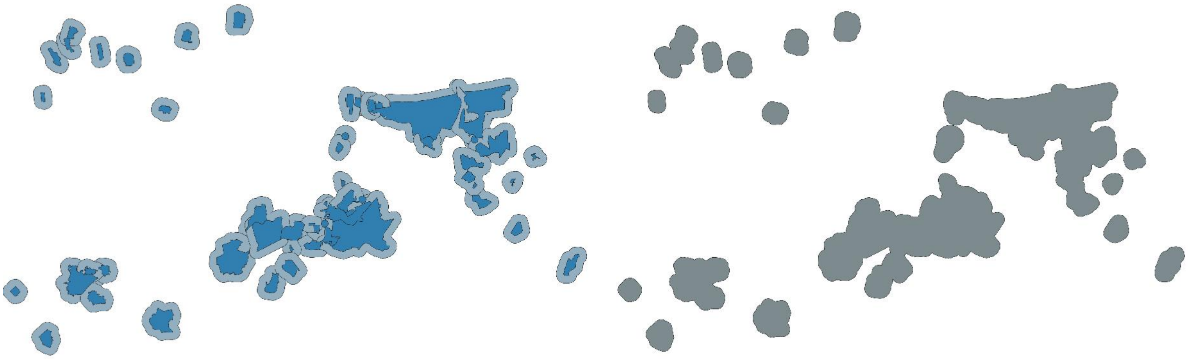
Appendices

1.A Appendix: Data

Constant geography

Geographical units covered by the Economic Census of Mexico are States, Municipalities, Localities, and AGEBs, in descending order. To capture the change of economic activities within a single region over time, we needed to generate an identifier to overcome the issue of localities growing in size and splitting into multiple localities.[27] Therefore we developed a balanced panel of agglomerations, which we generate by combining neighboring localities that share borders. First, we take the 2019 Economic Census as a baseline considering that it will have the most extensive coverage of localities. The geographical coverage of the Economic Census is based on economic activity, hence a combination of both urban and rural localities. The next step was then to merge both the urban and rural localities that appeared in the Census into the shapefiles published by the INEGI. However, in cases where we were not able to find a shapefile for a locality in the Census, we found an alternative source of the Catalog, also published by INEGI which is a list of localities and their coordinates. We transformed the list of coordinates into points on the map and created a 1km buffer around those points in order to factor them in as polygons. With the selected set of localities' polygons, we create a buffer of 1km to identify clusters of localities. If the buffered localities share borders, we define it as an "Agglomeration". This process yielded a total of 3,248 unique agglomeration IDs.

Figure 1.12: Constructing agglomerations



(a) 1km Buffer around localities (b) Grouped neighboring localities

Notes: Figure shows an arbitrary area.

[27] Localities are defined as 2,500 inhabitants.

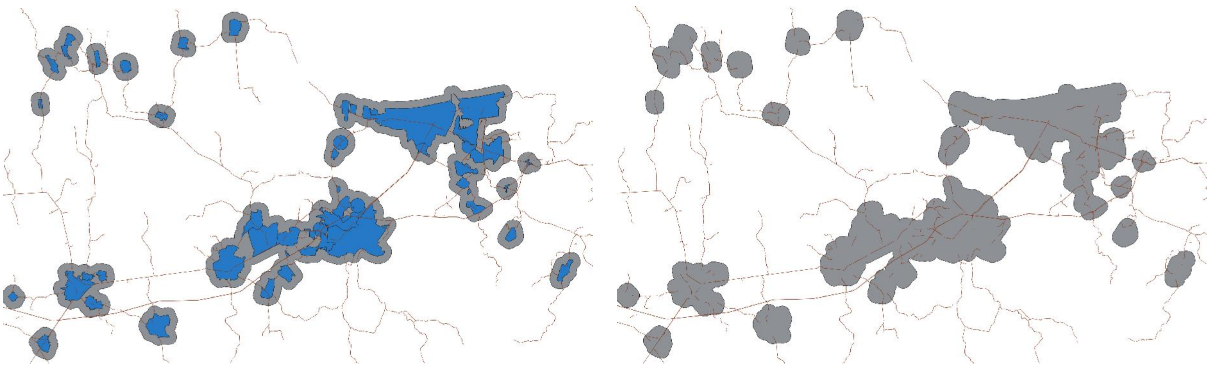
Once we had the polygon shapefile of agglomeration IDs, we assigned each year of localities in the respective Economic Census to a respective agglomeration id. This process is conducted in three steps. First, we repeat the process of selecting from the map which localities are covered by the Census. We then overlap the localities shapefile with the agglomeration shapefile to assign the ID of its overlapping agglomeration. For the localities that were not matched to the Economic Census, the second source was the Catalog, and for those that still did not find a correspondence, we assigned the same agglomeration id as the largest locality in the given municipality.

Roads

We use 2004, 2011, 2014, and 2019 highways (*Red Nacional de Caminos*) publicly available from INEGI. Among the different types of road constructions, we focus on intercity highways (Carreteras). Similar to agglomerations, we follow the assumption that highways cannot disappear. Therefore, we fixed the set of highways from each year by adding highways that existed in the previous year but were omitted. First, we created a buffer of 500m around all the highway maps to accommodate inconsistent breaks of highways. Next, we overlap the previous year's map with the more recent map, identifying which segments lie outside the buffer zones of the most recent map. This will locate the highways that exist in the previous year but not on the most recent map. Based on the assumption that highways do not disappear, we append these parts to the recent map and create a "fixed" map of highways.

The information available for each highway are the ID number of the highway, route number, speed, and the number of lanes.

Figure 1.13: Highways in 2018



(a) 1km Buffer around localities　　(b) Grouped neighboring localities

Notes: Figure shows an arbitrary area.

Table 1.13: National investment in highways (Million 2018 MXN)

	2004-2014	2014-2018
Total investment	599650.42	221466.46
Yearly avg.	59965.04	55366.61

Investment numbers taken again from the annuals. We take half the reported investment for end-of-period and start-of-period years.
Yearly values deflated with the inflation reported between July of the base year and July of 2018. Inflation taken from INEGI's inflation calculator available at https://www.inegi.org.mx/app/indicesdeprecios/calculadorainflacion.aspx.

Minimum travel times

Similar to grid points, but in more accurate distance measurement, we use Uber's Hexagonal Hierarchical Spatial Index as our grid system, or in other words, cells.[28] Based on the hexagonal system, we locate agglomerations and highway networks to the overlapping hexagon and store the highway's information in the respective hexagons. For instance, information on the speed, number of lanes, and width of the road will be stored in the neighboring hexagons, which allows us to develop an algorithm to estimate travel times that follow the path of hexagons. In addition to the highway information publicly available at INEGI, we consider the elevation of localities to reflect actual travel times.[29]

Our travel time analysis is conducted in four steps. First, we select the origin, destination, and highway network that will be used to travel from one location to another. It can be traveling from one agglomeration to another, or it could also be from one agglomeration to an airport, port, or even a specific city. For the origin and destination shapefiles, in case they are in polygons, we extract each polygon's centroids and consider them as a starting point and an ending point. Once we have chosen the shapefiles, we use the aforementioned open-source hexagon system by Uber to locate the shapefiles into respective hexagons. When storing information to the highways, we assign a set of parameters to address the issue of missing information for some years. We acknowledge that some highway shapefiles might not have all the information on speed, lanes, and width; hence we include in the algorithm to take specific values when there is a piece of missing information. Additionally, there will always be hexabins where it is not close to a highway network. For these hexagons, we assign a speed value of 5km/hour, meaning the only option will be to travel by walking. We also assign the order of variables based on priorities among speed, lane, and width for the code to first use when calculating the travel time. Then, we estimate the travel time from one origin point to all other destinations using the properties. Finally, we merge all the different origin points into a single matrix.

Construction plans

We focus on two government infrastructure projects under two administrations: Felipe Calderón(2006-2012) (see Figure 1.14) and Enrique Peña Nieto(2012-2018) (see Figure 1.15). Based on the official report National Infrastructure Program published by the Department of Transportation (Sector Comunicaciones y Transportes(SCT)), we focused on highway plans, which yielded

[28] https://www.uber.com/blog/h3/
[29] https://portal.opentopography.org/datasetMetadata?otCollectionID=OT.042013.4326.1

175 plans from the Calderón administration and 76 plans from the Peña Nieto administration. Both reports include details on which State the highway is located in, and the type of improvement the plan aims to achieve (construction or expansion).

Figure 1.14: Example of 2007-2012 Construction Plan

Carreteras Región Noroeste

Nombre / descripción	Entidad federativa	Monto total de inversión (miles de millones de pesos)	Fuente / esquema de financiamiento	Fecha de realización	
				Inicio	Término
Caborca-Sonoyta-San Luis Río Colorado-Mexicali					
Caborca-Sonoyta Ampliación a 12 metros (143.1 km)	Sonora	1.2	PEF	2006	2010
Sonoyta-San Luis Río Colorado Ampliación a 12 metros (192 km)	Sonora	1.4	PEF	2008	2011
San Luis Río Colorado-Mexicali Ampliación a 4 carriles (56 km)	Baja California	1.7	PEF	2006	2008
Ciudad Obregón-Hermosillo-Nogales					
Libramiento de Ciudad Obregón Construcción a 12 metros (45 km)	Sonora	0.7	Aprovechamiento de activos	2010	2011
Estación Don-Nogales Ampliación a 4 carriles (468.5 km)	Sonora	2.5	Aprovechamiento de activos	2010	2011
Libramiento de Hermosillo Construcción a 12 metros (37 km)	Sonora	0.9	Aprovechamiento de activos	2010	2011
Transpeninsular de Baja California					
Maneadero-Punta Colonet Ampliación a 12 metros (105 km)	Baja California	0.5	PEF	2009	2010
La Purísima-San Ignacio Ampliación a 12 metros (180 km)	Baja California Sur	2.0	PEF	2009	2012

Carreteras Región Noroeste

Nombre / descripción	Entidad federativa	Monto total de inversión (miles de millones de pesos)	Fuente / esquema de financiamiento	Fecha de realización	
				Inicio	Término
La Paz-Los Cabos					
Puentes paralelos El Piojito en La Paz Construcción a 4 carriles	Baja California Sur	0.1	PEF	2007	2008
La Paz-San Pedro Construcción a 4 carriles (15.5 km)	Baja California Sur	0.3	PEF	2007	2008
San Pedro-Todos Santos Ampliación a 12 metros (52 km)	Baja California Sur	0.5	PEF	2008	2009
Libramiento de Todos Santos Construcción a 12 metros (10 km)	Baja California Sur	0.3	PEF	2008	2009
Todos Santos-Cabo San Lucas Ampliación a 12 metros (73 km)	Baja California Sur	1.2	PEF	2007	2010
Mexicali-Laguna de Chapala					
Mexicali-San Felipe Ampliación a 12 metros y 4 carriles (150 km)	Baja California	1.1	PEF	2009	2012
Puertecitos-Laguna de Chapala Ampliación a 7 metros (110 km)	Baja California	0.6	PEF	2008	2010
Mazatlán-Culiacán					
Libramiento de Mazatlán Construcción a 12 metros (31 km)	Sinaloa	1.0	Aprovechamiento de activos	2009	2010
Libramiento de Culiacán Construcción a 12 metros (22 km)	Sinaloa	0.7	Aprovechamiento de activos	2009	2010

Figure 1.15: Example of 2013-2018 Construction Plan

Entidad Nombre de la obra	Inversión (MP)	Meta (km)	Tipo de Trabajo
SONORA			
Caborca-Sonoyta	66.8	8.6	Ampliación
Santa Ana-Altar 1/	0.0	1.7	Ampliación
Blvd. Álvaro Obregón en la Cd. Nogales	4.5	0.0	Ampliación
Puente Cabullona	5.0	0.0	Ampliación

Based on the construction plans, we develop a data set that contains information on the respective state in which the plan takes place, the duration of the project, and specific details of the construction.

Figure 1.16: Using Google Earth Pro to digitize the construction plans

Once we had a dataset of the construction plans list, the next step was to digitize the information in map format. Using Google Earth, we set the origin and destination of the highway plan based on how the name is written (e.g., if the plan stated "Expansion a 12 m Caborca-Sonoyta", meaning expand the lane to 12m in the highway connecting Caborca and Sonoyta, we would set Caborca as the origin and Sonoyta as the destination). Each search was saved, merged, then exported into a shapefile. However, note that the plans did not mention which specific part of the highway they will improve. Thus we considered the entire highway as a part of the plan. Once we had a complete shapefile of all the construction plans, we conducted a quality check for all the plans. We would search the plan online and see if there are additional sources published by each State government supplementing the details of the plan. In some cases, the State government reported an image of the exact location of the plan.

We were able to classify the construction plans based on the type of road, the type of improvement, and whether the targeted highways are located in/out of a city.

Type of highways listed in the construction plan:

- Inter-region (e.g., Chalco-Nepantla)
- Beltways (e.g., Libramiento)
- Connection to the border of each state (e.g., Límite de estados Pue/Ver)
- Junctions (e.g., entronque La Ventosa)

- Access to a specific location (e.g.,Acceso al Puerto Salina Cruz)
- Bridges (e.g., Puente)

Type of construction plans:

- Expand highways to 4/6/8 lanes (both direction)
- Construct 4 lanes (both directions)
- Expand or construct 2 lanes and 2 side roads
- Modernize and improve conditions

City classification:

- IN: If the construction plan is for a highway inside a city
- OUT: If the construction plan is for a highway connecting two regions outside a city
- LIB: If the construction plan is for beltways specifically [30]

Collecting information on whether the construction plan was executed.

Treatment variables

Variables For all the treatment variables, we generate three types of buffers around each agglomerations in order to accommodate the noise of map accuracy. All variables are constructed by agglomeration IDs.

Length of highways' segments that overlap each agglomeration Area of buffered agglomerations Density (length/area) of highways' segments

With regard to the construction plan, we first construct a dummy variable indicating whether an agglomeration lies within any construction plan. We assess by 5,10, and 15km buffers of each agglomeration. Next, we specify the construction plan by those that were executed and those that were not. We generate a dummy variable indicating whether an agglomeration lies within an executed construction plan and a non-executed plan. Lastly, we develop a dummy variable indicating whether an agglomeration is placed in a construction plan's starting and end points. We identify the starting and end point by the region's first and last name mentioned in the plan.

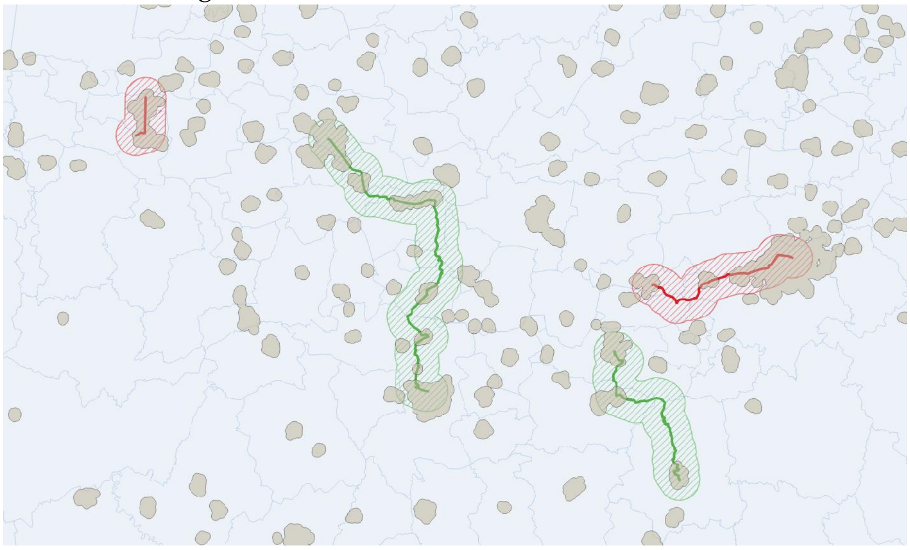
We also construct treatment variables to measure market access. Using the population census of 2019, we extract the population size by agglomerations. Next, we identify the top 100 agglomerations with the largest population. Then we use the previously generated minimum travel time values to generate a new variable, the distance from each agglomeration to the nearest hub.

[30] We specify the beltways since beltways mostly have the purpose of reducing traffic within each city.

1.B Appendix: Empirics

Treatment and sample

Figure 1.17: Treated and untreated locations



Notes: Figure displays an arbitrary area of the country. Gray areas are locations (3,248 in total). Dashed areas are buffers around construction plans. For a given year, green construction plans have been fully executed and red not yet.

Table 1.14: Treated locations

(a) Locations by overlap with plans

	2007-2012			2013-2018		
Buffer size (km)	5	10	15	5	10	15
With plans	771	1,052	1,330	457	678	898
With out plans	2,475	2,194	1,916	2,789	2,568	2,348
Total	3,246	3,246	3,246	3,246	3,246	3,246

(b) Locations by execution of plans

	2007-2012			2013-2018		
Buffer size (km)	5	10	15	5	10	15
Executed	259	261	265	278	404	551
Not executed	512	791	1,065	179	274	347
Total	771	1,052	1,330	457	678	898

Table 1.15: Firms in the sample and treated group

(a) Sample of firms

	2007-2012			2013-2018			
Buffer size (km)	5	10	15	5	10	15	Total
1998	2.09	2.16	2.26	1.65	1.69	1.73	2.78
2003	2.23	2.30	2.42	1.75	1.79	1.84	2.98
2008	2.72	2.81	2.96	2.14	2.20	2.27	3.67
2013	3.06	3.17	3.36	2.43	2.50	2.58	4.17
2018	3.43	3.56	3.78	2.73	2.81	2.92	4.74

(b) Treated firms

	2007-2012			2013-2018			
Buffer size (km)	5	10	15	5	10	15	Total
1998	1.43	1.50	1.52	0.73	0.76	0.80	2.78
2003	1.51	1.59	1.61	0.77	0.80	0.84	2.98
2008	1.82	1.92	1.95	0.97	1.01	1.07	3.67
2013	2.04	2.16	2.20	1.12	1.16	1.23	4.17
2018	2.26	2.40	2.45	1.26	1.31	1.40	4.74

Validity of empirical approach

Table 1.16: Predicting construction plans 2013-2018

	(1) Plan	(2) Execution
log(population)	0.0483***	-0.0262
	(0.00987)	(0.0173)
log(value added/workers)	0.0563***	0.0483
	(0.0138)	(0.0251)
Δ log(population)	0.384***	0.144
	(0.0566)	(0.104)
Δ log(value added/workers)	-0.0335*	-0.0123
	(0.0147)	(0.0287)
log(votes for PRI)	-0.0191*	-0.0191
	(0.00828)	(0.0149)
Observations	2146	611
R-sq	0.255	0.379

Notes: Standard errors are in brackets. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$. Regressions at the municipality level. Variables and growth rates are from Economic Census 2003 and 2008 and population Census 2000 and 2010.

Table 1.17: Balance table

Variable	Locat.	Mean untreated	Mean treated	Diff.	s.e.	p-value	Stat. signif.
Share manuf.	400	0.169	0.151	-0.018	0.015	0.220	
Share salaried	400	0.228	0.288	0.061	0.019	0.001	***
log(K)	400	11.237	11.547	0.310	0.310	0.318	
log(K/L)	400	4.305	4.527	0.223	0.109	0.041	**
Δ share manuf.	376	0.018	0.013	-0.006	0.007	0.407	
Δ share salaried	376	-0.019	-0.025	-0.006	0.012	0.612	
Δ K	376	0.441	0.445	0.004	0.074	0.957	
Δ K/L	376	-0.001	0.072	0.073	0.067	0.275	
log(L per estab.)	400	0.944	0.996	0.052	0.049	0.289	
log(V.A./L)	400	4.665	4.859	0.194	0.138	0.161	
log(TFP) (L-P)	400	4.028	4.258	0.229	0.128	0.073	*
Δ L per estab.	376	0.114	0.082	-0.032	0.027	0.229	
Δ V.A./L	376	-0.202	-0.260	-0.058	0.065	0.373	
Δ TFP	376	-0.177	-0.170	0.007	0.094	0.937	
log(population)	397	10.949	10.964	0.015	0.174	0.931	
Δ population	362	0.801	0.832	0.030	0.015	0.050	*
log(highways)	400	10.753	10.618	-0.136	0.093	0.147	
Δ highways	400	0.271	0.236	-0.036	0.023	0.114	

First stage regressions

Construction plans and market access. An implicit assumption of our identification strategy is that the execution of construction plans affects firms by increasing their market access as they can reach more distant markets or acquire intermediate inputs at a lower cost. We test this assumption by estimating the following two-ways-fixed-effects model:

$$log(MA_j) = time + treatment_j + \delta \cdot time \cdot treatment_j + \beta \cdot controls_j + \varepsilon_j \tag{1.45}$$

We estimate Equation 1.45 separately for both sets of construction plans at the location level. Here $time$ denotes pre and post treatment periods and $treatment_j$ whether the location belongs to the treatment group or not. $controls_j$ is a battery of locaiton level controls at baseline.

MA_i is a measure of market access. We follow Allen and Arkolakis (2014); Blankespoor et al. (2017) to compute it according to:

$$log(MA_i) = \sum_j \frac{Population_j}{\tau_{i,j}^{\sigma-1}} \tag{1.46}$$

To stay consistent with the literature, we assume $\sigma = 9$. MA_i captures the market access from location i, defined as the weighted sum of the population of all locations in the country discounted by the one-to-one trade costs τ_{ij}. For this exercise, we keep the population fixed at 2003 levels. We compute two versions of this measure. MA_1 that includes all locations; and MA_2, that includes all but the location i itself. The trade costs τ_{ij} is determined as in Equation 1.30, explained in detail in the model section.

Table 1.18: First stage regressions

Plans 2007-2012			Plans 2013-2018		
	(1) $log(MA_1)$	(2) $log(MA_2)$		(3) $log(MA_1)$	(4) $log(MA_2)$
time	0.00752***	0.00812***	time	0.00130***	0.00154***
	(0.00180)	(0.00163)		(0.000188)	(0.000125)
treated	-0.00439*	-0.00383	treated	-0.000180	-0.000397
	(0.00245)	(0.00236)		(0.000625)	(0.000576)
time*treated	0.00798*	0.00757*	time*treated	0.000756**	0.000560*
	(0.00448)	(0.00444)		(0.000357)	(0.000329)
Controls	Yes	Yes	Controls	Yes	Yes
Obs.	1230	1230	Obs.	750	750
R-sq	0.99	0.99	R-sq	0.99	0.99

Notes: Standard errors are in brackets. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table 1.18 shows the results for both sets of construction plans, for two measures of market access and controlling for baseline characteristics market access in 2004 and state fixed effects. In summary, execution of construction plans has a positive effect on market access in treated locations.

For the construction plans 2007-2012, their execution implied a 0.79% higher market access for exposed locations. In this period, market access increased in average 0.75% for all locations, meaning that the treatment implied an increase in market access twice as large for treated locations. For the 2013-2018 plans, the increase was 0.07%. Since in this period market access increased in average 0.13% for all locations, the treatment implied a 53% larger market increase for treated locations.

Robustness checks

Regressions by sector

Table 1.19: Regressions by sector, Construction plans 2007-2012

Manufacturing

	(1) $log(va/L)$	(2) $log(TFP)_{GNR}$	(3) $log(L)$	(4) $log(w)$	(5) $Entry$	(6) $Exit$
β_{-1}	.0733*	0.01	0.0226	0.0062	0.0084	.0203*
s.e.	[.0401]	[.0148]	[.022]	[.0107]	[.0117]	[.0121]
β_0	-.0463*	0.0091	-.0559***	-0.0056	0.008	-.0199***
s.e.	[.0237]	[.0066]	[.0161]	[.0064]	[.0092]	[.0059]
β_1	.0648*	0.0129	-.0555*	-0.0107	0.0099	-.0337***
s.e.	[.0342]	[.0144]	[.0287]	[.0073]	[.0134]	[.0082]
Controls	No	No	No	No	No	No
Obs.	733,654	733,654	733,654	733,654	733,654	733,654

Commerce

	(1) $log(va/L)$	(2) $log(TFP)_{GNR}$	(3) $log(L)$	(4) $log(w)$	(5) $Entry$	(6) $Exit$
β_{-1}	-0.0372	-0.014	-0.0029	-0.0054	0.0099	0.0099
s.e.	[.023]	[.0158]	[.0071]	[.006]	[.0062]	[.0084]
β_0	0.0178	0.0092	-0.0116	-0.0045	0.0048	-.0144**
s.e.	[.0118]	[.0078]	[.0082]	[.0057]	[.0059]	[.0067]
β_1	.0835***	.0361**	-0.0055	-.0153**	0.0056	-.029**
s.e.	[.0235]	[.016]	[.009]	[.0073]	[.0049]	[.0137]
Controls	No	No	No	No	No	No
Obs.	2,727,356	2,727,356	2,727,356	2,727,356	2,727,356	2,727,356

Services

	(1) $log(va/L)$	(2) $log(TFP)_{GNR}$	(3) $log(L)$	(4) $log(w)$	(5) $Entry$	(6) $Exit$
β_{-1}	-0.03	-0.0081	-0.004	-0.0069	0.0082	0.003
s.e.	[.0258]	[.0097]	[.0104]	[.0064]	[.0089]	[.0086]
β_0	-0.0015	0.0038	-.017**	-0.0013	0.021	-.0203**
s.e.	[.021]	[.0045]	[.0069]	[.0058]	[.0141]	[.0096]
β_1	.0533***	.0188**	-0.0019	-.0108*	.0217***	-.0388***
s.e.	[.0202]	[.0095]	[.0123]	[.0058]	[.0075]	[.0143]
Controls	No	No	No	No	No	No
Obs.	3,511,463	3,511,463	3,511,463	3,511,463	3,511,463	3,511,463

Notes: Standard errors are in brackets. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table 1.20: Regressions by sector, Construction plans 2013-2018

Manufacturing

	(1) $log(va/L)$	(2) $log(TFP)_{GNR}$	(3) $log(L)$	(4) $log(w)$	(5) $Entry$
β_{-1}	-0.0505	-0.0127	0.0344	0.0036	0.0114
s.e.	[.0478]	[.0158]	[.0219]	[.0068]	[.0085]
β_0	0.0351	0.0092	0.0214	.0199**	-0.0009
s.e.	[.048]	[.0156]	[.0319]	[.0079]	[.0109]
β_1	-.1506**	-.0628***	0.0378	-0.0128	0.029
s.e.	[.0587]	[.024]	[.0263]	[.012]	[.0241]
Controls	No	No	No	No	No
Obs.	645,540	645,540	645,540	645,540	645,540

Commerce

	(1) $log(va/L)$	(2) $log(TFP)_{GNR}$	(3) $log(L)$	(4) $log(w)$	(5) $Entry$
β_{-1}	-.0562**	-.0182*	-0.0084	0.0048	0.009
s.e.	[.0222]	[.0102]	[.0056]	[.0035]	[.0057]
β_0	.0362**	.0173*	.0198*	.0081**	-0.0018
s.e.	[.0158]	[.01]	[.0102]	[.0036]	[.0076]
β_1	-.0438*	-0.0107	.0343***	0.0088	0.0171
s.e.	[.0247]	[.0186]	[.0088]	[.0131]	[.0105]
Controls	No	No	No	No	No
Obs.	2,521,552	2,521,552	2,521,552	2,521,552	2,521,552

Services

	(1) $log(va/L)$	(2) $log(TFP)_{GNR}$	(3) $log(L)$	(4) $log(w)$	(5) $Entry$
β_{-1}	-.0652**	-.0182**	-0.0067	0.002	0.0074
s.e.	[.0298]	[.0084]	[.0103]	[.0056]	[.0065]
β_0	.078**	.0207**	0.0111	.0098*	0.0049
s.e.	[.0312]	[.0099]	[.007]	[.0059]	[.0082]
β_1	.0811***	.0209***	.0337***	0.0231	.0287**
s.e.	[.0233]	[.0073]	[.0129]	[.0145]	[.0135]
Controls	No	No	No	No	No
Obs.	3,144,586	3,144,586	3,144,586	3,144,586	3,144,586

Notes: Standard errors are in brackets. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Regressions with controls

Table 1.21: Regressions with controls. Construction plans 2007-2012

	(1) log(va/L)	(2) log(TFP) (GNR)	(3) log(L)	(4) log(w)	(5) Entry	(6) Exit
β_{-1}	-0.025	-0.0097	-0.0003	-0.0054	0.0097	0.0074
s.e.	[.0209]	[.0102]	[.0078]	[.0044]	[.0069]	[.0071]
β_0	0.0095	.009**	-.0143***	-0.0007	0.0141	-.0173**
s.e.	[.0125]	[.0044]	[.0054]	[.0038]	[.0102]	[.0072]
β_1	.0748***	.0295***	-0.0021	-.0106**	.0152***	-.0329**
s.e.	[.0216]	[.011]	[.0069]	[.0045]	[.005]	[.0131]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	7,060,649	7,060,649	7,060,649	7,060,649	7,060,649	7,060,649

Notes: Standard errors are in brackets. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$. Sample includes all firms from 1998 to 2018. Excludes firms with value added or capital smaller than zero. Controls include 3-digit sector.

Table 1.22: Regressions with controls. Construction plans 2013-2018

	(1) log(va/L)	(2) log(TFP) (GNR)	(3) log(L)	(4) log(w)	(5) Entry
β_{-1}	-.0629***	-.0193***	-0.0062	0.0037	0.0099
s.e.	[.0239]	[.0072]	[.0078]	[.0046]	[.0066]
β_0	.0494**	.0156**	.0134*	.0089**	0.0018
s.e.	[.0205]	[.0069]	[.0079]	[.0038]	[.0076]
β_1	-0.0033	-0.005	.0258***	.0129*	.025**
s.e.	[.0219]	[.0095]	[.0094]	[.0069]	[.0125]
Controls	Yes	Yes	Yes	Yes	Yes
Obs.	6,375,668	6,375,668	6,375,668	6,375,668	6,375,668

Notes: Standard errors are in brackets. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$. Sample includes all firms from 1998 to 2018. Excludes firms with value added or capital smaller than zero. Controls include 3-digit sector.

Regressions by buffer size

Table 1.23: Buffer = 10km, Construction plans 2007-2012

	(1) log(va/L)	(2) log(TFP) (GNR)	(3) log(L)	(4) log(w)	(5) Entry	(6) Exit
β_{-1}	-.0414*	-0.0177	-0.0059	-0.0082	.0118*	0.0003
s.e.	[.0217]	[.0123]	[.008]	[.0051]	[.0063]	[.0072]
β_0	0.0059	0.0092	-0.0119	-0.0014	0.0148	-.0138*
s.e.	[.0209]	[.0077]	[.0074]	[.0053]	[.0092]	[.0079]
β_1	.0598***	.0152*	-0.0059	-.0132**	.0191***	-.0334***
s.e.	[.0211]	[.0088]	[.0095]	[.0053]	[.0043]	[.0128]
Controls	No	No	No	No	No	No
Obs.	7,280,866	7,280,866	7,280,866	7,280,866	7,280,866	7,280,866

Notes: Standard errors are in brackets. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table 1.24: Buffer = 10km, Construction plans 2013-2018

	(1) log(va/L)	(2) log(TFP) (GNR)	(3) log(L)	(4) log(w)	(5) Entry
β_{-1}	-.065**	-0.0078	-0.004	0.0024	0.0078
s.e.	[.0265]	[.0076]	[.0074]	[.0049]	[.0054]
β_0	.0564**	.0184***	.0154*	.0108**	0.0017
s.e.	[.0221]	[.0071]	[.0089]	[.0048]	[.0077]
β_1	0.0116	0.0133	.0372***	.0148**	.0277**
s.e.	[.0241]	[.0105]	[.0101]	[.0062]	[.0119]
Controls	No	No	No	No	No
Obs.	6,526,519	6,526,519	6,526,519	6,526,519	6,526,519

Notes: Standard errors are in brackets. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table 1.25: Buffer = 15km, Construction plans 2007-2012

	(1) log(va/L)	(2) log(TFP) (GNR)	(3) log(L)	(4) log(w)	(5) Entry	(6) Exit
β_{-1}	-0.0351	-0.0143	-0.004	-0.0075	.0111*	0.0038
s.e.	[.0217]	[.012]	[.0076]	[.0047]	[.0061]	[.0072]
β_0	0.0013	0.0083	-0.0109	-0.0007	.0161*	-0.0117
s.e.	[.0191]	[.007]	[.0077]	[.0051]	[.0088]	[.0076]
β_1	.0522**	0.0138	-0.002	-.0126**	.0216***	-.0284**
s.e.	[.0211]	[.0084]	[.0098]	[.005]	[.0045]	[.0132]
Controls	No	No	No	No	No	No
Obs.	7,665,879	7,665,879	7,665,879	7,665,879	7,665,879	7,665,879

Notes: Standard errors are in brackets. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table 1.26: Buffer = 15km, Construction plans 2013-2018

	(1) log(va/L)	(2) log(TFP) (GNR)	(3) log(L)	(4) log(w)	(5) Entry
β_{-1}	-.0718***	-0.0089	-0.0042	0.0036	0.0076
s.e.	[.0255]	[.0075]	[.0076]	[.0048]	[.0053]
β_0	.0611***	.0183***	.015*	.0096**	0.0023
s.e.	[.0217]	[.007]	[.0084]	[.0047]	[.0075]
β_1	0.0137	0.0131	.0361***	.0147**	.029**
s.e.	[.0238]	[.0102]	[.0098]	[.0061]	[.0117]
Controls	No	No	No	No	No
Obs.	6,723,947	6,723,947	6,723,947	6,723,947	6,723,947

Notes: Standard errors are in brackets. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Evidence on firm mobility, 2013-2018

In this section we show a novel margin of firm dynamics that can be affected by the development of the highways network: the geographical location of firms within cities. First, we show that firm mobility is present in the data by exploiting a novel section in the Economic Census 2018 where establishments are asked if they had a different location in the previous wave (2013), and then report the main reason why they moved. And second, by regressing the mobility decision and reasons on execution of construction plans.

Firm mobility in the data. The 2018 census included two new questions regarding firm mobility. First, the census asks if the firm changed address between 2013 and 2018. If the answer is yes, the census asks an additional question on the reason why it moved to a different address. The reasons are codified in 6 categories: low business growth, increase in facility's rental prices, to move closer to clients and suppliers, public insecurity, tax-related reasons and, finally, other reasons.

Not all firms answered the firm mobility questions. The 2018 census covers 4,737,931 firms; among them, 1,832,685 answered the mobility questions, which is 39% of the total. According to INEGI's officials, small and medium firms are over-represented among the respondent firms. Considering only the respondents, 4.28% of firms changed address between 2013 and 2018, which means that the census documents 78,527 movers. By extrapolating this percentage to the full census, the number of movers could be around 203,011 firms. However, this number could be biased if non-respondents have a different moving behavior.

Among the 78,527 movers, 12.6% are from the manufacturing sector, 32.6% from commerce, and 54.8% from services. In the population of firms, 12.3% are in the manufacturing sector, 47.6% in commerce, and 40.1% in services. If moving to a different location was random, we should expect these percentages to be similar. However, there is a large disparity in the share of movers from the services sector and the share they represent in the population. This suggests that service providers are more likely to move to another location. A possible explanation could be that they face lower moving costs or expect higher returns from moving than firms in commerce and manufacturing.

Firms might have many reasons to move. The Economic Census asks what is the main one and codifies the answers. The distribution of these answers is the following. 10.43% declare low business growth, 31.8% increase in facility's rental prices, 13.8% to move closer to clients and suppliers, 3.6% public insecurity, 0.8% tax-related reasons and, finally, 39.5% other reasons.

The effects of better highways on firm mobility. We now provide evidence on the effects of highways on the firm mobility decision. To do this, we estimate the following probit model:

$$P(\text{new location in 2018 = yes})_{ij} = \Phi[\alpha + \beta \mathbf{X}_{ij} + \delta D_j + \varepsilon_{ij}] \tag{1.47}$$

In this model, i denotes the firm and j the location. $\mathbf{X}_{ij}$ denotes a vector of controls, and D_j takes the value of 1 if construction plans were executed between 2013 and 2018 and zero otherwise. The parameter of interest is δ, which captures whether better highways affect the probability of moving to a different location.

Table 1.27 shows the results by sector and with and without controls for population density and number of firms and workers at baseline, to control for the fact that mobility might defer depending on how crowded a location is. Columns (1) and (2) show that execution of contraction plans has a positive effect on the probability of an firm to have moved to a different location between census waves of 2013 and 2018. Columns (3) and (4) show that manufacturing firm mobility doesn't seem to be connected to changes in the highways network. Finally, firms in the services sector seem to be affected by highways when they make mobility decisions but these effects are not robust to baseline demographic characteristics of the location.

Table 1.27: Porbit model. Outcome: probability of moving

Sector	**Commerce**		**Manufacturing**		**Services**	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	.1269**	.0731**	0.0907	0.0699	.0992**	0.0595
se	-0.0484	-0.0349	-0.0552	-0.0552	-0.0497	-0.04
N	475,370	472,852	124,885	124,464	476,515	476,031
Controls	No	Yes	No	Yes	No	Yes

Notes: Standard errors are in brackets. $^{***}p < 0.01,^{**}p < 0.05,^{*}p < 0.1$.

Similarly, to determine if highways affect the reasons why firms move, we run the following probit model:

$$P(\text{main reason } r = \text{yes})_{ij} = \Phi[\alpha + \beta \mathbf{X}_{ij} + \delta D_j + \varepsilon_{ij}] \tag{1.48}$$

Where r is the main reason why the firm changed location and can be: low business growth, increase in facility's rental prices, to move closer to clients and suppliers, public insecurity, tax-related reasons and, finally, other reasons. Table 1.28 shows the results by sector and adding controls for baseline demographic characteristics such as population density, number of firms and workers. Whereas firms can move for diverse reasons, when highways are improved, the reported reason that is positively distorted is proximity to clients and suppliers, except for the manufacturing sector.

Table 1.28: Porbit model. Outcome: probability of moving

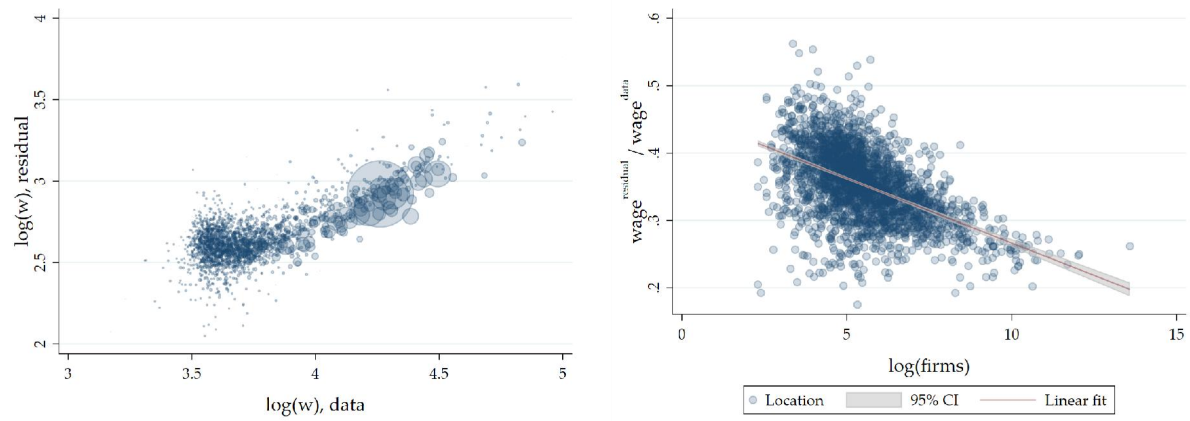
			Commerce			
	(1)	(2)	(3)	(4)	(5)	(6)
Reason	Growth	Rents	Proximity	Insecurity	Taxes	Other
Treatment	-0.0045	-0.0693	.1113**	0.077	0.1018	-0.0097
se	[.0376]	[.0632]	[.0434]	[.0648]	[.0705]	[.0543]
N	14,672	14,672	14,672	14,672	14,672	14,672
Controls	Yes	Yes	Yes	Yes	Yes	Yes
			Manufacturing			
	(1)	(2)	(3)	(4)	(5)	(6)
Reason	Growth	Rents	Proximity	Insecurity	Taxes	Other
Treatment	-0.0023	-0.006	0.0392	0.0226	.3036**	-0.021
se	[.0498]	[.0714]	[.0558]	[.0704]	[.112]	[.072]
N	5,802	5,802	5,802	5,802	5,802	5,802
Controls	Yes	Yes	Yes	Yes	Yes	Yes
			Services			
	(1)	(2)	(3)	(4)	(5)	(6)
Reason	Growth	Rents	Proximity	Insecurity	Taxes	Other
Treatment	0.0007	-0.0771	.1241***	0.0441	0.086	-0.0073
se	[.0271]	[.0489]	[.028]	[.0439]	[.0685]	[.0444]
N	25,220	25,220	25,220	25,220	25,220	25,220
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are in brackets. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

The impact of new highways on intra-city relocation choices carries an implication for firm dynamics. When location changes are not tracked in the data, the rate of firm exit can be inflated, possibly leading to an underestimation of the reduction in exit observed produced by our treatment. Simultaneously, not tracking location changes could lead to an overestimation of firm entry which could potentially result in an overestimation of firm entry rates in treated locations. Lastly, considering that relocations are often motivated by a desire to be closer to clients and suppliers, it is reasonable to expect that revenue productivity tends to be higher at the new locations.

1.C Appendix: Model

Figure 1.18: Residual wage by location

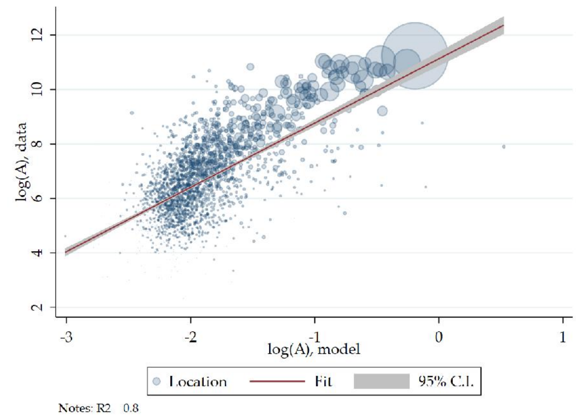


(a) Wage by location, data and residual

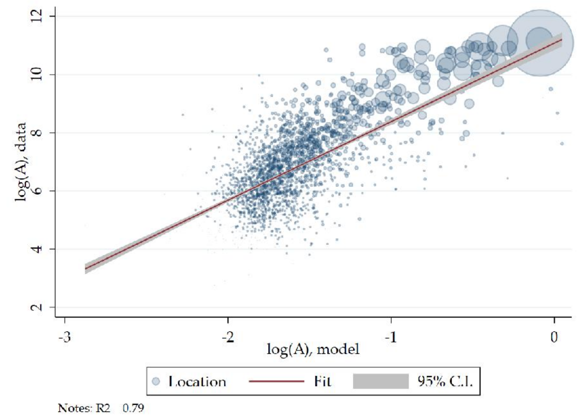
(b) Residual wage/data wage by location

Notes: Figure shows estimation for 2018. Marker size in panel (a) denotes the number of firms; the largest is Mexico City. $\beta_1, \beta_2, \beta_3$ significant at the 95%.

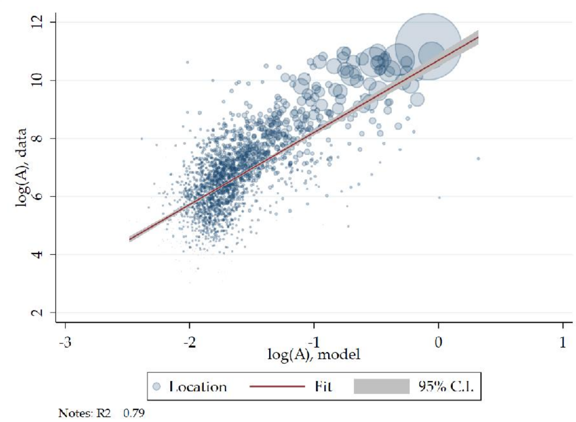
Figure 1.19: Local labor productivity, model vs. microdata



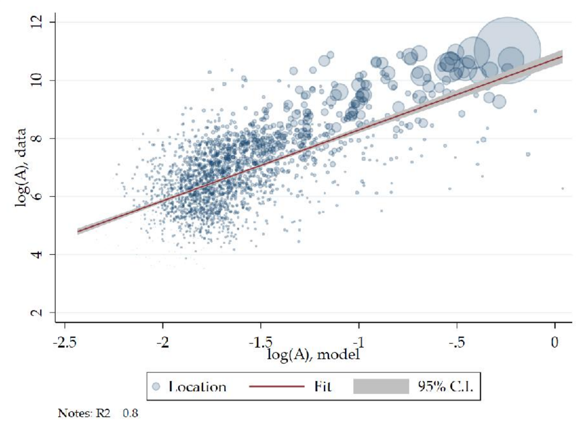
(a) 1998



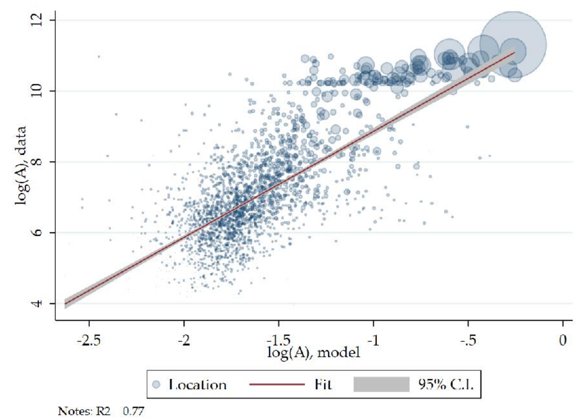
(b) 2003



(c) 2008



(d) 2013

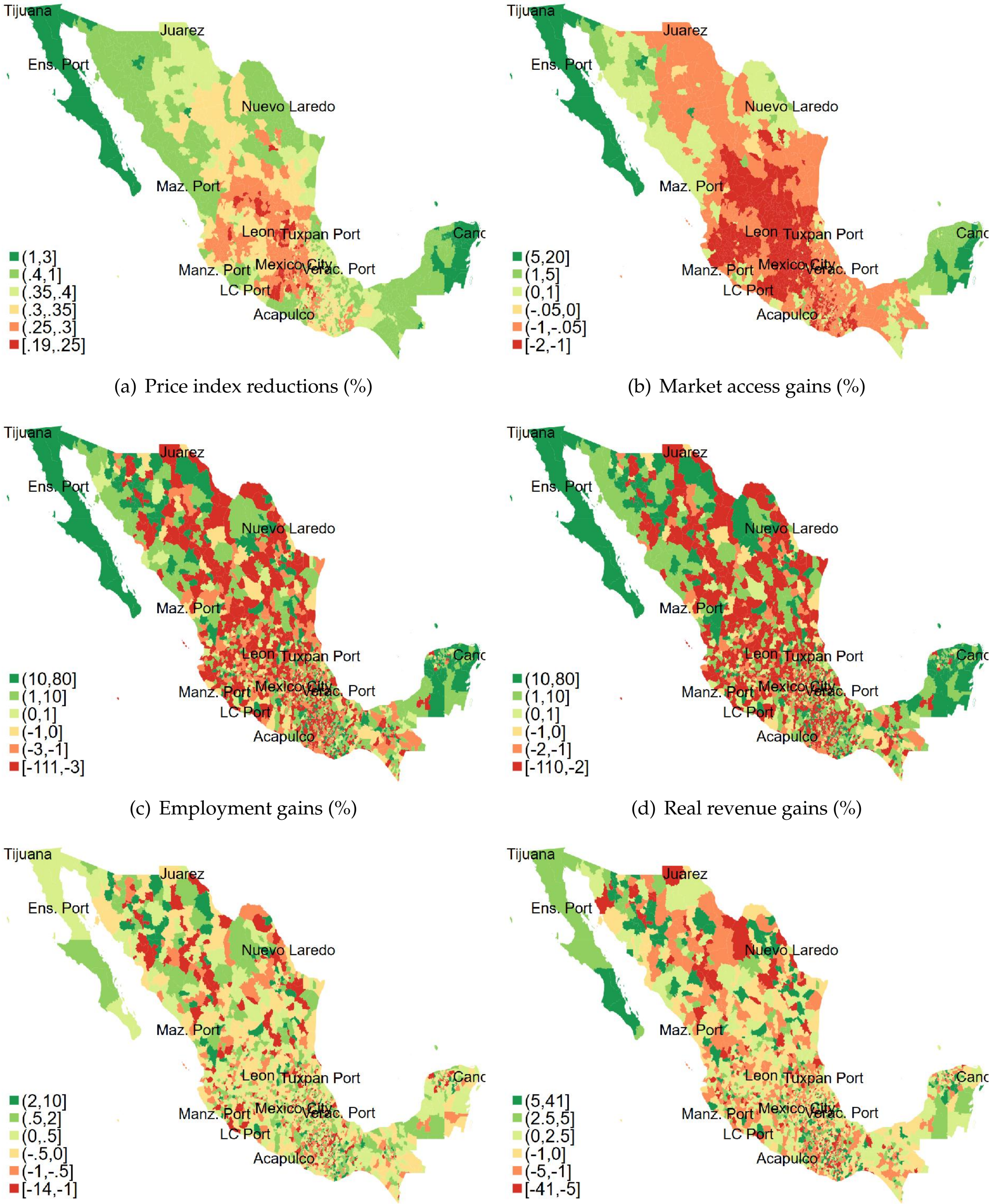


(e) 2018

Notes: Gains stemming from expanding the highways network from 1998 to 2018. Gains are at the location level.

Quantitative results

Figure 1.20: Location gains from the 1998-2018 highways network



(a) Price index reductions (%) (b) Market access gains (%)

(c) Employment gains (%) (d) Real revenue gains (%)

(e) Productivity gains (%) (f) Firms gains (%)

Notes: Gains stemming from expanding the highways network from 1998 to 2018. Gains are at the location level.

1.C.1 Agglomeration and congestion externalities

Agglomeration externalities. Firm level productivity is separable in two parts as:

$$\psi_{i,t}(n) \equiv z_{i,t} \cdot s_i(n) \tag{1.49}$$

Now $z_{i,t}$ is not fully exogenous but depends positively on the local population to capture agglomeration externalities stemming from, for example, a larger pool of ideas that make all workers more productive in the location.

$$z_{i,t} = \bar{z}_{i,t} L^{\alpha}_{i,tz} \tag{1.50}$$

Where $\bar{z}_{i,t}$ is the exogenous part and $\alpha_z \geq 0$ governs the degree of agglomeration externalities. The rest of the model remains the same. The existence of the spatial equilibrium will now depend on α_z. Allen and Arkolakis (2014) provide the existence conditions.

Congestion externalities. Utility is still given by:

$$U_{i,t} \equiv C_{i,t} \cdot u_{i,t} \tag{1.51}$$

But now, local amenities suffer from congestion externalities. The larger the amount of people living in a location, the larger the degradation and congestion of amenities. We can model it as:

$$u_{i,t} = \bar{z}_{i,t} L^{\alpha}_{i,tu} \tag{1.52}$$

Where $\bar{u}_{i,t}$ is the exogenous part and $\alpha_u \geq 0$ governs the degree of congestion externalities. Adding a congestion force reduces the strong negative relationship between local wages and amenities. Allen and Arkolakis (2014) provide the existence conditions of the equilibrium for combinations of parameters governing agglomeration and congestion externalities.

1.C.2 Firm sorting

We model firm sorting following Gaubert (2018). Firms choose their location only at entry since, in the data, most firm migration happens within locations rather than across them.[31] Productivity $\psi_i(n)$ of firm n when choosing location i is:

$$\log(\psi_{i,t}(n)) = \log(z_{i,t}) + \alpha \log(L_{i,t}) + \log(s_i(n)) \cdot (1 + \log(L_{i,t}))^{\eta} + \varepsilon_{i,t}(n) \tag{1.53}$$

Here, $z_{i,t}$ captures location-specific productivity shocks, α governs the intensity of local spillovers, η determines the degree of complementarity between idiosyncratic productivity $s_i(n)$ and city size, and finally, $\varepsilon_{i,t}$ represents a firm-level taste shock for location i.

Equation 1.53 fully determines the sorting of firms. Local spillovers rationalize why more firms move to big cities, the complementarity term explains why big cities attract the most productive firms, and the taste shock accounts for the imperfect sorting observed in the data.

We assume that the taste shock follows a Fréchet distribution with shape parameter ξ. The

[31] Alternatively, this empirical observation can be replicated by allowing on-the-life-cycle migration coupled with high moving costs.

share firms that will choose location i is:

$$\frac{M_{i,t}}{M_t} = \frac{V_{i,t}^{\xi}}{\sum_{j \in \mathcal{J}} V_{j,t}^{\xi}} \tag{1.54}$$

Chapter 2

Making a Growth Miracle: Historical Persistence and the Dynamics of Development

Oscar Fentanes[1] & Jonas Gathen[2]

Abstract

What explains growth miracles? We argue that growth miracles are driven by a fundamental race: as the economy tries to catch-up to its steady state, changes in the economic environment move the steady state itself and provide new potential for catch-up growth. We quantify this race over the course of development using 40 years of plant-level manufacturing panel data from Indonesia and a structural model of plant dynamics. We estimate the model on the micro data along the observed growth path without assuming that the economy is ever at a steady state. While catch-up growth starting from initial conditions in 1975 accounts for 42% of Indonesia's subsequent industrialization, new changes in the economy induce new catch-up growth. In the end, the economy never catches up.

JEL Codes: D25, O11, O14, O47

Keywords: Development, industrialization, growth miracles

[1]Toulouse School of Economics; oscar.fentanes@gmail.com.

[2]Toulouse School of Economics; j.gathen@posteo.de. We are grateful for the continued support and many useful comments by our advisors: Fabrice Collard, Matteo Bobba, Christian Hellwig and Stéphane Straub. We further thank in no particular order: Ben Moll, Yongseok Shin, Francisco Buera, Ezra Oberfield, Richard Rogerson, Jihyun Kim, Eugenia Gonzalez-Aguado, Tuuli Vanhapelto, Christophe Bruneel-Zupanc, Miguel Zerecero, Alexandre Gaillard, Andreas Schaab, Gyung-Mo Kim and all participants of the TSE Macro workshop. We want to thank the World Bank Jakarta team, especially Massimiliano Calí, for gratefully sharing the manufacturing plant-level data for Indonesia. We further thank feedback during presentations at Princeton, WashU St Louis, EWMES 2022, the Annual STEG conference 2022, EEA/ESEM 2022, ENTER jamboree 2021 as well as the PhD and Macro workshops at TSE and LSE. All remaining errors are our own.

2.1 Introduction

Over the past 50 years, extended periods of rapid economic growth in China, India and Indonesia alone lifted roughly 1 billion people out of extreme poverty (World Bank 2023). What drives such growth miracles? A common view is that growth miracles capture a transition process: a permanent policy regime change sets an initially poor and highly misallocated economy on a transition path towards a better long-run equilibrium (Buera and Shin 2013; Asturias et al. 2023). These transitions take time because frictions prevent labor and capital to quickly reallocate across firms and sectors. But if transitions take time, how does this view account for new changes in policies while the economy is still transitioning? And how can we identify the aggregate effects of new policies if the economy is still adjusting to previous policies?

In this paper, we show evidence that instead of a one-time transition, growth miracles are driven by a never-ending race: as the economy tries to catch-up to its steady state, changes in the economic environment move the steady state itself and induce new transition growth. We do so by looking at Indonesia, the fourth most populous country in the world. Indonesia provides an ideal case to study this race since we can draw on almost half a century of manufacturing plant-level panel data during which the Indonesian economy completely transformed: GDP per worker increased five-fold, the working population tripled, output in manufacturing grew 30-fold and the manufacturing employment share doubled.

Drawing on the micro data, we provide empirical evidence that motivate a model of plant dynamics in which growth is driven both by changes in the economic environment and transition phases. Empirically distinguishing these two drivers of growth is crucial because one might otherwise falsely attribute current growth to current policy changes. To do so, we estimate the model on the micro data along the observed growth path without assuming that the economy is at a steady state at any point in time. Intuitively, observed conditional choices of plants identify changes in the economic environment, while the distribution over these choices summarizes the past and reveals the potential for transition growth.

In line with the common view of growth miracles, we find that initial transition growth is important: letting the economy in 1975 transition while shutting down all future changes in the economic environment explains 42% of the manufacturing growth between 1975 and 2015. However, we also find that the Indonesian economy in 2015 is not closer to its steady state than it was in 1975, precisely because new changes in the economic environment in the meantime moved the steady state itself. As the economy is always undergoing important transition processes, one key implication is that evaluating policies without considering these adjustments is highly misleading.

We now provide further details and results. Drawing on our data, we document four main facts that help us disentangle transition growth from changes in the economic environment and motivate our subsequent model:

Fact 1: Rapid economic growth coincided with changes in the plant distribution. Average plant size doubled, the mass of plants increased four-fold and the right tail of the plant size distribution thickened.

Fact 2: Adjustment processes can account for changes in the plant distribution. These are respectively: An aging of the plant distribution together with the fact that plants enter small and grow over their life cycle, slow entry and exit dynamics, and the fact that it takes time to grow large plants.

Fact 3: The drivers of aggregate productivity differed markedly before and after the Asian Financial Crisis. Before the crisis, productivity across plants was driven entirely by the selection of more productive plants, informed by meager within-plant productivity growth. In contrast, the post-crisis growth period was characterized by strong within-plant productivity growth and little growth due to selection.

Fact 4: The allocation of resources did not improve systematically over time. This is robust to different measures of misallocation. We find evidence for volatile productivity dynamics at the plant-level, large changes in entry, and plant-level adjustment frictions – particularly in labor – from an event study design that can jointly account for this.

To quantify the race between transition growth and changes in the economic environment, we draw on a model of plant dynamics in the tradition of Hopenhayn (1992) that is motivated by the previous empirical facts. In the model, firms face risk regarding their productivity, choose to enter and exit and hire labor and capital subject to adjustment frictions that lead to the slow accumulation and reallocation of resources across sectors and firms. We embed these plant dynamics into a two-sector economy to capture the endogenous reallocation of workers across manufacturing and the rest of the economy. The main frictions in manufacturing are labor adjustment costs, in particular convex costs, that prevent plants from growing a large workforce quickly (*Facts 2 + 4*). Plants also endogenously enter and exit based on drawing entry costs and fixed costs of production. The level and dispersion of costs in turn rationalize the observed speed of entry and exit dynamics (*Facts 2*). These features imply that with an initial distribution characterized by few but productive young plants, the economy goes through a process of transition growth as plants gradually grow, more plants gradually enter over time and unproductive plants gradually exit. At any given point in time, the model economy is characterized by a set of exogenous model fundamentals and the state of the current economy as captured by the distribution of plants. Model fundamentals include all cost parameters as well as time-varying aggregates such as aggregate labor supply and technological changes in manufacturing (*Fact 3*) and the rest-of-the-economy. Policy affects growth through driving part of the changes in model fundamentals. Technically, we assume that plants make dynamic choices forming rational expectations over their future idiosyncratic risk but have perfect foresight over future aggregate changes in the economy. This introduces a computationally difficult fixed point problem: plants' dynamic choices depend on expectations over the future path of market-clearing prices, which in turn depend on the endogenous evolution of the entire distribution of plants (as in Buera and Shin 2013).

The key methodological contribution of the paper is to propose a tractable estimation strategy that allows to estimate this model economy on standard plant-level micro data along the growth path in the data without assuming that the observed economy is at a steady state at any point in time. Importantly, our model estimation allows model fundamentals to vary flexibly over time, making it particularly suited to study fast-changing economies and markets. The model estimation proceeds in three main steps that allow to distinguish transition growth from changes in fundamentals and make the computational costs of the estimation independent of the computational costs of solving for a path of model equilibria. In the first step, we identify the path of time-varying equilibrium prices – only wages in our case – along the observed growth path (e.g. as in: Gopinath et al. 2017). Given that our model can account for this equilibrium path, we can treat the path of wages as fixed throughout the estimation and thereby avoid solving for the computationally costly fixed point in the path of equilibria. In the second estimation step, we identify the distribution of plants over the state space of the model, summarizing the history of the economy. In this step, we estimate plant production functions (Ackerberg, Caves, and Frazer 2015; Demirer 2020) and propose a novel strategy to separate plant-level productivity into an idiosyncratic and a common aggregate technology component, allowing us to distinguish selection-driven from technology-driven productivity growth. In the third and last estimation step, we estimate the model parameters that govern plants' adjustment frictions. We do so by drawing on Euler equation Continuous Conditional Choice (CCC) estimation, exploiting observed conditional input and exit choices of plants and avoiding to solve a dynamic programming problem to compute model-based dynamic input choices (Hotz and Miller 1993; Bajari, Benkard, and Levin 2007). In this last step, we estimate sizable convex adjustment costs in our model, which are identified from the empirical pattern that even small but highly productive plants grow their labor force gradually over time.

Using the estimated model, we find that transition growth from starting the economy with initial conditions in 1975 and shutting down all future changes in model fundamentals explains 42% of subsequent manufacturing output growth and all of the aggregate welfare increases that are due to changes in manufacturing over time. Given an initial distribution that features young and small plants, sizable labor adjustment frictions and slow entry and exit dynamics, it takes the economy 26 years to reach 90% of the steady state manufacturing output. Importantly, transition growth remains an important driver of growth precisely because the economy's fundamentals continue to change. To quantify this point, we repeat the previous exercise to compute the transition path for each year, starting from each year's initial distribution and model fundamentals. We find that the economy does not get closer to its (time-varying) steady state. It takes on average 20 years to reach 90% of the steady state manufacturing output and – if anything – the time it takes increases over time. Based on our results, we can thus strongly reject the idea that transition growth is a transitory phenomenon.

Large changes in fundamentals are key to explain the continuing importance of transition growth. The structural model allows us to quantify the role of changes in fundamentals in Indonesia's manufacturing growth miracle and quantify how much of their effect can be explained by changes in observed government policy. We do so by focusing on two important changes in fundamentals that can be linked to development policies that the Indonesian government also pursued to varying degrees over the 40 years we study: (1) large-scale investments in education that raise the pool of skilled (and cheap) labor, and (2) the active use of FDI policy to attract manufacturing plants under foreign ownership.

We find that the manufacturing growth miracle would not have happened in the absence of the estimated doubling in human capital per worker, because labor would have been more expensive in this economy and manufacturing plants are far more sensitive to higher wages than the rest of the economy. To gauge the importance of policy in driving overall human capital increases, we then evaluate Indonesia's largest school construction program (INPRES) through the lens of the model. Building on micro-empirical evidence on the wage effects (Duflo 2001, 2004), the scale of the program (Akresh, Halim, and Kleemans 2023), and the slow labor market integration of treated cohorts, we show that by 2015, the program accounts for roughly

10% of the overall manufacturing output growth that is due to human capital per worker increases in the economy.

In contrast, for FDI, we find that manufacturing output in 2015 would have only been 8% lower in the complete absence of foreign-owned entrants, while we find that regulatory changes in FDI policy in the late 1980s may potentially account for 85% of the overall effect of FDI on manufacturing. Taken together, this still means that most growth by far stems from structural forces related to demographics. Thus, a somber conclusion based on these results – partly resonating related work on the Indian growth miracle (Bollard, Klenow, and Sharma 2013) – is that policy matters less for growth than we might think.

Related literature

We contribute to four main strands of literature. First and most importantly, we complement the growth and macro development literature by studying a model of firm dynamics where growth is driven by the combination of changes in exogenous fundamentals and transition growth. This is in contrast to a firm dynamics literature that has mostly analyzed development differences through differences in steady states.[3] Much fewer papers study transition growth with firm dynamics (e.g. Buera and Shin 2013; Moll 2014; Akcigit, Alp, and Peters 2021; Ruggieri 2022; Asturias et al. 2023; Lanteri, Medina, and Tan 2023). We add to this literature by (1) allowing for further changes in fundamentals along the transition and by (2) estimating this model on the micro data along the transition without assuming that the economy is at a steady state at any point in time. Quantitatively, we find that the combination of both sources of growth matters. Not only are initial transition growth and changes in fundamentals important for growth, but the economy is always far away from its (time-varying) steady state, questioning the usefulness of either comparing steady states or focussing on transitions in the absence of further changes in fundamentals. Apart from the methodological differences, Buera and Shin (2013) and Asturias et al. (2023) are the most closely related in their focus on understanding growth miracles. Our results mainly differ from Buera and Shin (2013) in that we find no role for reductions in frictions and misallocation – their main driver of transition growth – but rather a key role for plant and worker demographics in driving transition growth. We return to Asturias et al. (2023) below.

Second, both modeling and key results in this paper relate to the recent quantitative spatial, trade and migration literature, which focus on frictional worker mobility and trade while abstracting from firm dynamics. For example, the idea that transitions take long and the economy is persistently far away from its steady state resonates with recent findings from Allen and Donaldson (2020) and Kleinman, Liu, and Redding (2023). This literature relates to and builds on the seminal work of Caliendo, Dvorkin, and Parro (2019) who also study the combination of changes in exogenous fundamentals and transition growth. In contrast to Caliendo, Dvorkin, and Parro (2019), tractability in our case does not come from dynamic hat algebra techniques and we estimate all time-varying model fundamentals. Slow transitions in our paper and this literature share common causes: low and highly dispersed exit (moving) probabilities and slow input adjustments.

[3]While the literature on firm dynamics with a development focus is too vast to cite, overviews are for example given by Hopenhayn (2014) and Restuccia and Rogerson (2017) for misallocation, Ulyssea (2020) for informality and Alessandria, Arkolakis, and Ruhl (2021) for trade.

Third, we contribute more generally to the Quantitative Macroeconomics literature by showing how to tractably estimate the model directly on the observed transition path in the data, identify the model entirely on plant-level data and only use macro moments for model validation, an approach that we see as closely aligned to a growing literature that moves "from micro to macro" (see the overview in: Buera, Kaboski, and Townsend 2023). The "equilibrium estimation" methods that ensure tractable estimation – enforcing the observed path of equilibria throughout the estimation and Euler CCC estimation – are used in other literatures, but have not yet seen wider application in the Macroeconomic literature.[4] We find the equilibrium estimation approach to be particularly suited for studying a path of time-varying equilibria, since we can also tractably estimate entire paths of time-varying parameters. While estimation methods that require to first solve the model may offer more flexibility on the choice of moments that identify parameters, they often have to strongly restrict the parameter space.

At last, our paper also relates to the literature on growth and productivity dynamics. Our results of selection-driven aggregate productivity growth in Indonesian manufacturing mirrors similar results in Brandt, Van Biesebroeck, and Zhang (2012) for China and Asturias et al. (2023) for Chile and Korea. We add to these papers by establishing the result of selection-driven productivity growth in a non-parametric setting that nests a larger class of growth models including various endogenous growth models.

The rest of the paper is structured as follows. The next section presents the main empirical evidence. Section 2.3 develops the model and discusses identification, estimation and model validation. In Section 2.4, we quantify the main drivers of growth. The last section concludes.

2.2 Empirical evidence

In this section, we introduce the Indonesian data and key facts about the Indonesian growth experience that motivate the subsequent model.

2.2.1 Data

Our primary data comes from the plant-level Annual Manufacturing Survey, collected by Indonesia's Central Bureau of Statistics. It covers only medium- to large-sized manufacturing plants by surveying all formal manufacturing establishments with more than 20 employees. The survey contains detailed and consistent annual information on standard plant-level characteristics from 1975 to 2015, a period of 41 years. It covers between roughly 7,500 to 30,000 plants per year. Throughout, we draw on reported information on plants' age (based on birth year), industry (up to 5-digit) and ownership (including foreign ownership). On the production side, we draw on plants' capital stock, value-added revenue, and the number of workers (including paid and unpaid workers) and wage bill (including contributions and in-kind compensation) which are separately reported for production and non-production workers. Unfortunately, capital is only reported starting in 1990. All variables denoted in Indonesian Rupiah

[4]The idea of estimating models on the observed path of equilibria in Macroeconomics dates at least back to Hansen and Singleton (1982). More recent research that conditions estimation on the observed path of equilibrium prices can for example be found in the literature following Caliendo, Dvorkin, and Parro (2019). In a recent working paper, Humlum (2022) exploits these two estimation steps in a similar general model framework of growth and firm dynamics, although entirely different context of industrial robot adoption in Denmark.

are deflated to real values using the aggregate CPI and normalized to the year 2010. For aggregate data, we further combine the GGDC 10-sector and Economic Transformation databases for Indonesia, which capture time-consistent aggregate sectoral employment and output series over the time period 1960-2012 (Timmer, de Vries, and De Vries 2015) and 1990-2018 (Kruse et al. 2023) respectively. We refer to this data as the *GGDC* series throughout. In Appendix 2.A.1, we discuss in detail the data cleaning and homogenization steps we take to ensure consistency of all datasets over time.

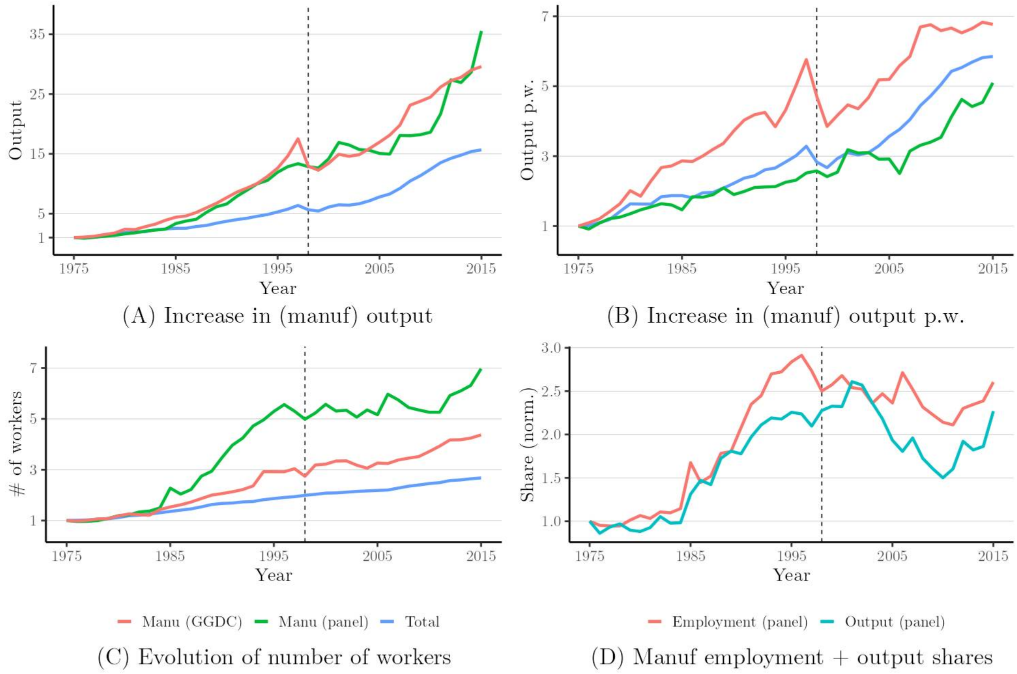
There are two important data limitations. First, while the data is in principle a census of plants with 20 or more workers, in practice the census misses plants. This shows up in discontinuous jumps in new plant entry during years of the economy-wide Economic Census in 1975, 1985, 1995 and 2006. In those years, plants are added that were previously missed, either because they were small and made the cutoff, or newly entered. For the subsequent analyses, this means that aggregate changes often show discontinuous drops in census years and should actually look more smoothed out. Given that we observe the census in 1975, the initial distribution is correctly reported. Furthermore, this does not bias results that are based on within-plant variation. Our data also misses plants because of non-reporting, either because plants miss to report in some years or because we are forced to drop a plant-year entry due to misreporting (see Appendix 2.A.1). We treat these missing entries as missing at random and specifically account for missing entries in our structural model. We correct our measurements of plant entry and exit by denoting plant entry as the first time when a plant identifier enters the panel and plant exit if we do not observe a given plant identifier at any future time period (see: Appendix 2.A.1).

The second main data limitation is with respect to the coverage of our data. As we show in Appendix 2.A.2, while our dataset misses the approximately 99% of Indonesian manufacturing plants that have less than 20 workers, most of these plants are characterized by self-employment with a modal plant size of one to two workers. After cleaning, our manufacturing data captures between 25-30% of total manufacturing employment and value-added output as based on the *GGDC* data, with shares increasing over time (Figure A.1). We think of small scale manufacturing as a separate sector given robust evidence that there are few transitions between small and larger scale manufacturing (e.g. Poschke 2013; Van Biesebroeck 2005; Schoar 2010) and most plants that enter our panel have only recently been established. For example, the median age of plants that newly enter our plant panel is only two years. The focus in this paper is thus on how relatively large plants and their dynamics drive aggregate economic growth. In the model and results parts, we explicitly model the entire economy, taking into account that our data only captures a time-varying share of output in the overall economy.

2.2.2 Four key facts of growth

We now highlight four key facts that shed light on the Indonesian growth experience. The first two facts relate to changes in the entire plant distribution and the importance of slow adjustment processes. Fact 3 and 4 document changes in the drivers of productivity growth and the absence of improvements in the allocation of resources over time.

Figure 2.1: Evolution of aggregate and sectoral employment and output



(A) Increase in (manuf) output

(B) Increase in (manuf) output p.w.

(C) Evolution of number of workers

(D) Manuf employment + output shares

Notes: (Economy-wide) Total gives the aggregate of the GGDC data. Panel refers to the Indonesian manufacturing plant census (1975-2015, 20+ workers). All series are normalized by their respective value in the first year. (A) and (B) use value-added output. Dashed vertical lines denote the onset of the Asian Financial Crisis in 1997.

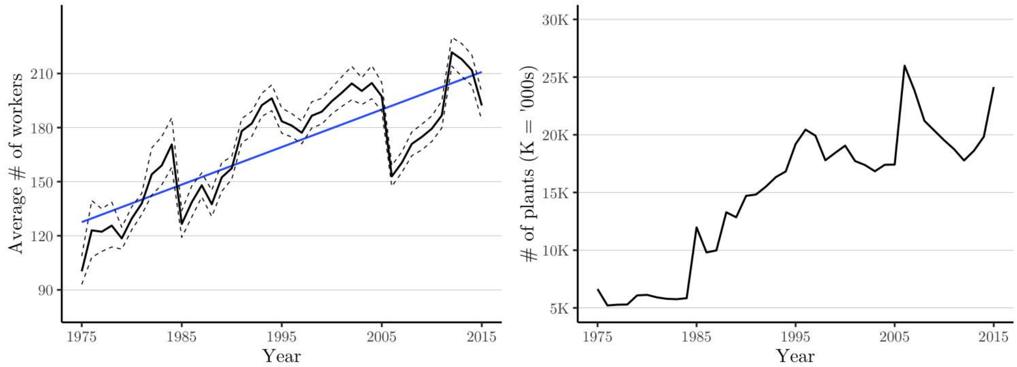
Rapid economic growth coincided with changes in the plant distribution

As the 4th most populous country in the world, Indonesia underwent a dramatic process of economic development over the past 50 years, a few key features of which are reported in Figure 2.1. Between 1975-2015, GDP per worker increased more than five-fold (Panel B), driven by a 15-fold increase in output (Panel A) and roughly a tripling of the working population (Panel C). Manufacturing contributed importantly to this aggregate growth: output grew 30-fold and the manufacturing employment share more than doubled (Panel D).

The period 1975-2015 can be divided into two main growth regimes that are separated by the Asian Financial Crisis in 1997. The pre-crisis period captures a period of rapid labor-intensive industrialization, including the period 1987-1994 that Hausmann, Pritchett, and Rodrik (2005) characterize as a growth acceleration. Most of the total worker flows into manufacturing happen before 1997 and manufacturing grows far more rapidly than the rest of the economy. Fast growth in the aggregate working population is also a defining feature of the pre-crisis period with an average annual growth rate of 3%, 70% higher than in the post-crisis period. Based on our census of medium- and large-sized plants, the rise of manufacturing rapidly takes off in the first half of the 1980s and industrialization peaks with the Asian Financial Crisis as evidenced by manufacturing employment and output shares (Panel D). After the Asian Financial Crisis, which started in 1997 and hit manufacturing mostly in 1998, the economy experienced lower total output growth that was due to substantially lower growth in plant entry and employment and – as we will show further below – by higher plant-level productivity growth.

These aggregate changes went in hand with systematic changes among manufacturing plants. As evidenced in Figure 2.2, the rapid increase in the total number of workers in manufacturing

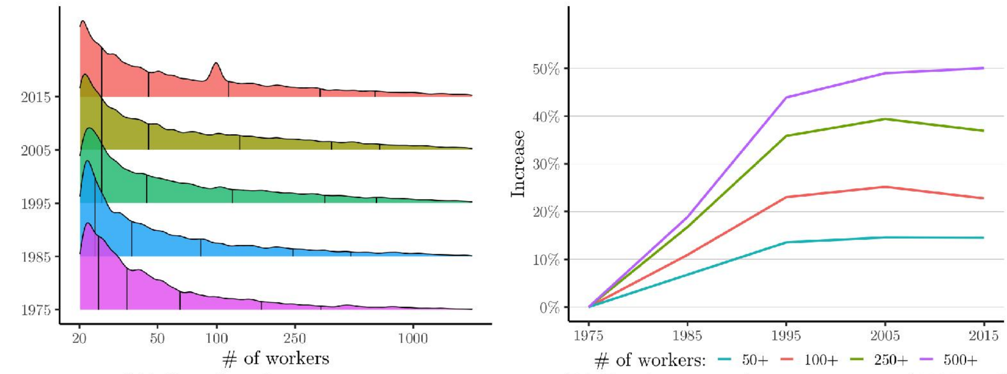
Figure 2.2: Evolution of average plant size and number of plants



(A) Avg number of workers over time (B) Number of plants over time

Notes: Data based on Indonesian manufacturing plant census (1975-2015) with 20+ workers. Panel A: Workers include paid and unpaid workers. Dotted lines give 95% bootstrapped confidence intervals and solid blue line gives best linear fit. Panel B: Jumps in 1985, 1995 & 2006 are explained by Economic Census years.

Figure 2.3: Evolution of employment distribution over time



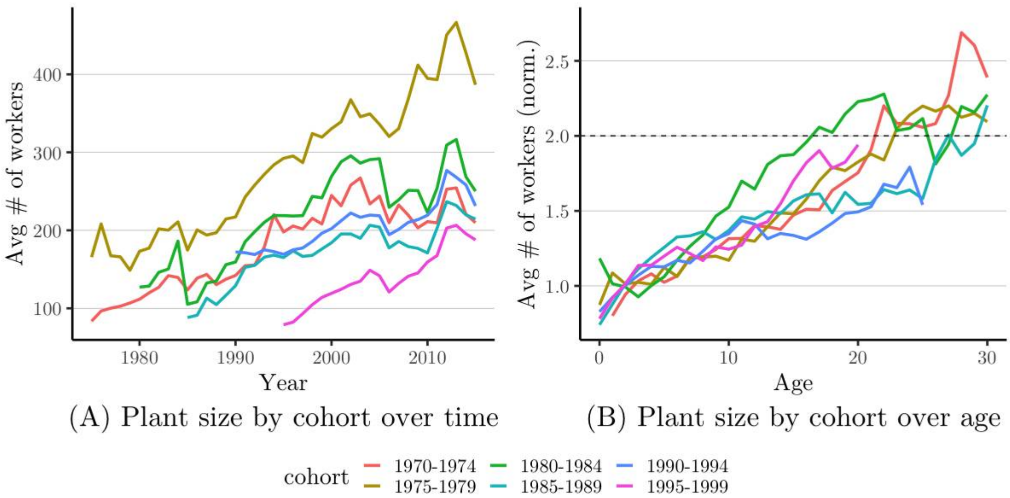
(A) Size distribution over time (B) Employment shares over time (1975 = 1)

Notes: Data based on Indonesian manufacturing plant census (1975-2015) with 20+ workers. Only showing years 1975, 1985, 1995, 2005 and 2015. Panel A: Vertical lines give 25th, 50th, 75th, 90th and 95th percentiles. Bunching at 99 workers starts in 2013, after the passing of the 2012 Worker Safety Law that binds for establishments with 100+ workers. Panel B: Share of employment in plants with more than X workers.

is met with a doubling of the average number of workers in manufacturing plants (Panel A) and a four-fold increase in the number of manufacturing plants. In both cases, most of the gains were already reached by 1997. Importantly, the entire plant distribution changed systematically over time. Specifically, Figure 2.3 shows that the right tail of the plant employment distribution systematically thickened over time – a key feature of the development process that has been highlighted for firms (rather than establishments) across and within countries (Chen 2022; Choi et al. 2023; Poschke 2018). Panel B shows that the employment share in plants with more than 50 workers increased by roughly 15%, while the employment share in plants with more than 500 workers increased by more than 50%.[5] The increase in the right tail of the employment distribution is a main driver of the increase in the average plant size over time.

[5]We report this metric as it is a simple transformation of the Pareto tail, which is also robust to left-censured data. We report the corresponding secular decline in Pareto tail coefficients in Appendix 2.A.2. We find systematic changes in the Pareto coefficient, both in the cross-section (which is not in line with a common Pareto distribution) and over time (which is not in line with traveling wave equilibria).

Figure 2.4: Plant life cycle growth by birth cohort



(A) Plant size by cohort over time

(B) Plant size by cohort over age

Notes: Data based on Indonesian manufacturing plant census (1975-2015) with 20+ workers. Workers include paid and unpaid workers. Plant age is based on reported year of establishment. Panel B normalizes each entry by the cohort-specific average plant size of plants below age 5 (as in Hsieh & Klenow 2014). Note that each cohort over time is an unbalanced panel as only surviving plants stay in the panel and there is (limited) plant entry as plants make the cutoff of 20 workers to be included in the census.

Adjustment processes can account for changes in the plant distribution

We now show evidence for three slow adjustment processes that can respectively account for increases in the plant size, the mass of plants and the right tail of the plant size distribution.

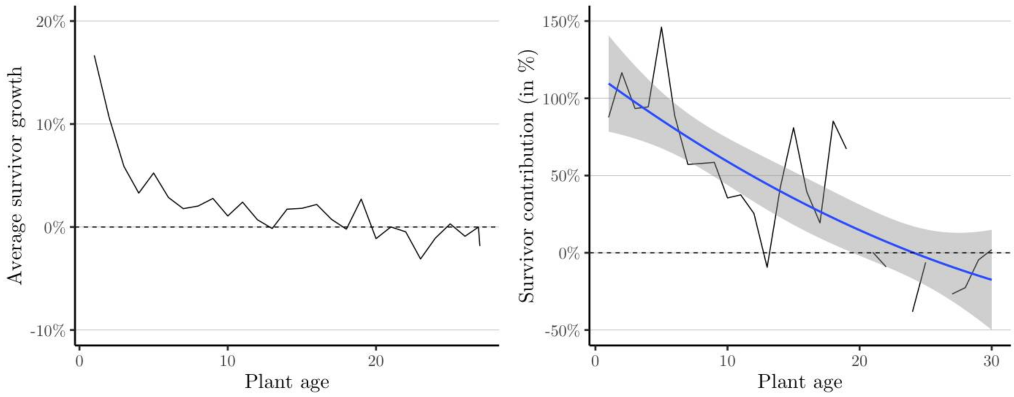
The aging of the plant distribution We start by showing that the slow aging of the plant distribution – average plant age increased by 40% since 1975 – can explain average plant size increases. The reason is that plants enter small and grow over their life cycle. Figure 2.4 plots life-cycle growth profiles across different cohorts of surviving manufacturing plants. Plants enter roughly with a similar average number of workers, which grows with plant age. Plants that survive at least 20 years have about twice as many workers as new entrants; in comparison and as documented in Hsieh and Klenow (2014), manufacturing plants in the US that survive for that long are about six times as large as new entrants.[6]

A benefit of our panel data is that we can show that the increase in the average size of surviving plants is mostly driven by within-plant growth rather than selection (larger plants being more likely to survive). Figure 2.5 Panel A shows average year-to-year within-plant growth by age. Young plants grow their employment quickly, with growth declining slowly as plants become older, average growth running out after around 20 years. Panel B reports the relative contribution of survivor growth to average plant size increases by age using the following accounting identity:

$$\underbrace{\bar{L}_a - \bar{L}_{a-1}}_{\Delta\text{avg plant size}} \equiv \underbrace{\frac{N_a^S}{N_a}(\bar{L}_a^S - \bar{L}_{a-1}^S)}_{\text{Survivor contribution}} + \underbrace{\frac{N_a^E}{N_a}\bar{L}_a^E}_{\text{Entry}} - \underbrace{\frac{N_{a-1}^X}{N_{a-1}}\bar{L}_{a-1}^X}_{\text{Exit}} + \underbrace{\bar{L}_{a-1}^S\left(\frac{1}{N_a} - \frac{1}{N_{a-1}}\right)}_{\text{Net reallocation}} \tag{2.1}$$

[6]The numbers are not perfectly comparable, because of the cutoff of 20 workers in the Indonesian data. If young plants in the US data are smaller, this overestimates the difference across countries, while plants that stay below 20 workers in the US data bias downward.

Figure 2.5: Within-plant (survivor) growth in employment



(A) Avg employment growth by age (B) Relative contribution of survivors

Notes: Data based on Indonesian manufacturing plant census (1975-2015) with 20+ workers. Sample restricted to survivors and plants for which plant age is defined (N = 543,586). Panel A: Average annual growth in total number of workers (paid + unpaid) across plants by age, weighted by a plant's previous employment. Panel B: Relative contribution of survivor growth to average plant size growth based on accounting identity in Equation 1 and dividing by the left-hand side to obtain relative contributions.

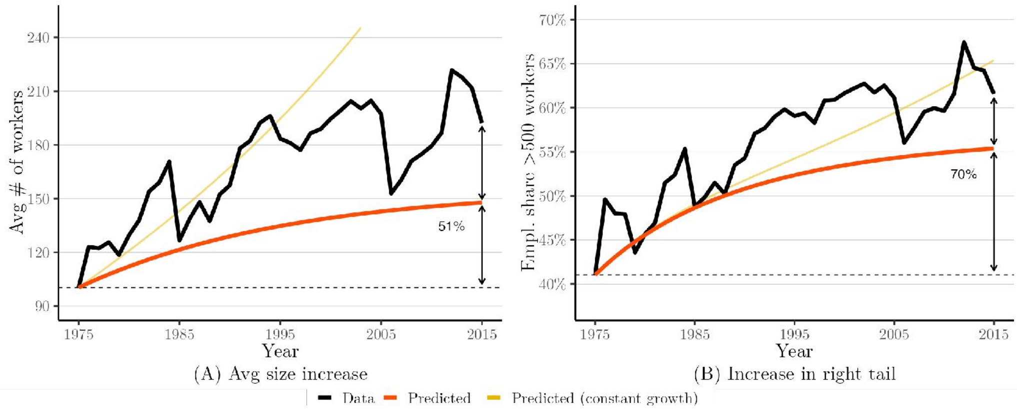
where S refers to the set of surviving plants from age $a-1$ to age a, E refers to entering plants (which exist because of the size threshold in our data) and X refers to exiting plants. The contribution of survivors, entry and exit respectively measure their average size weighted by their share in the population of plants. The net reallocation effect is driven by changes in the total number of plants over age: if exit outweighs entry (as in our case) then workers are reallocated towards fewer plants, mechanically increasing average plant size. We find that for young plants, growth by survivors explains all of the increase in average employment across plants, while selection as given by the remainder dominates the total effect after age 10-15.

Together, this evidence implies that average plant size crucially depends on where the age distribution of plants is; since the initial distribution of plants in the newly emerging manufacturing sector in 1975 featured mostly young plants, average plant size was small. Despite the rapid entry of new plants, plants became older and hence larger over time. For example, average plant age increased by 40% between 1975 to 2006 (the last year in which plant age was reported in the survey). Figure A.5 also shows how the entire age distribution shifted right over time.

It takes time for entry and exit dynamics to play out Next, we highlight a basic driver of slow entry and exit dynamics that can partially account for the observed four-fold increase in the mass of plants: low entry and exit rates. The idea is simple: the young manufacturing sector in 1975 features few plants and if only few potential entrepreneurs move into and out of entrepreneurship, it takes time to build up a mass of plants. For example, suppose there is a fixed mass of potential entrepreneurs and entry and exit rates into entrepreneurship are the same. Then the long-run (steady state) share of entrepreneurs is 1/2, but if the economy starts with no entrepreneurs, it can take many years to get close to the steady state. For entry and exit rates at 7.9% – equal to the average exit rate across all years and plants in our data[7] – it

[7] This exit rate is substantially lower than the 14-18% documented for informal establishments in Vietnam (McCaig and Pavcnik 2021) and slightly smaller than the 8.3% documented for small establishments across 12 developing countries (McKenzie and Paffhausen 2019). Within manufacturing, exit rates also seem to decline with plant

Figure 2.6: Reduced-form transition dynamics implied by initial plant size distribution



Notes: Predicted changes in distribution based on exercise taking discretized initial plant size distribution (# of workers) in 1975 and transition matrix giving conditional probabilities of moving from one plant size bin to another for years 1975-1976. Predictions iterate on initial discretized distribution with fixed transition matrix. Predicted (constant growth) instead enforces transition matrix incorrectly enforcing constant growth taking average plant size growth for 1975-1976.

already takes the economy 14 years to reach just 90% of the long-run steady state. While we do not generally observe potential entrants and thus cannot study entry rates without further assumptions, in the Appendix we show additional evidence for slow exit processes. Specifically, Figure A.8 shows that exit rates only slowly decrease with plant productivity and that exit rates do not increase with aggregate shocks such as the Asian Financial Crisis. This is in line with growing evidence that stagnant firms in developing countries tend to survive longer compared to firms in developed countries (e.g. Hsieh and Klenow 2014; Akcigit, Alp, and Peters 2021; Eslava, Haltiwanger, and Pinzon 2022).

It takes time to grow large plants At last, we show that a lack of large plants in the 1970s and the fact that it takes time to grow large plants can jointly explain the slow fattening of the right tail of the employment distribution. We do so by considering the following exercise. Take as the starting point the initial employment distribution of plants Φ_0 in 1975 and discretized in $X = 10$ different size bins. Each bin captures the fraction of plants with this number of workers. We now follow individual plants and compute the probability of moving from one bin to the other between 1975 and 1976, which we summarize in the transition matrix P of dimension X^2. We predict changes in the distribution by iterating on the initial distribution using the fixed transition matrix: $\hat{\Phi}_{t+\tau} = \Phi_0 \cdot P^\tau$. Figure 2.6 shows that the exercise explains 50% of average plant size increases (A) and 70% of the increases in the employment share of plants with more than 500 workers (B) over time. The reason is that in 1975, the distribution lacks large plants in comparison to the stationary distribution implied by the employment growth observed between 1975 and 1976 and it takes time to grow large plants. The exercise predicts that it takes 25 years to reach 90% of the steady state average plant size, broadly capturing the speed at which plant growth plays out over time.

The exercise is robust to a number of concerns which we address in Appendix 2.A.3.[8] Im-

size, explaining a lower exit rate of 6.2% across all US manufacturing (e.g. Clementi and Palazzo 2016) and broadly similar exit rates for all manufacturing plants in India and Mexico (Hsieh and Klenow 2014).

[8]Specifically, we show similar results when allowing for entry and exit, taking any other starting years in the 1970s and that the slow filling up of the right tail of the employment distribution holds for any other year-to-year

portantly, the exercise cannot distinguish the drivers of employment growth. The transition matrix $P_{t,t+1}$ only gives a reduced-form summary of plant employment growth subject to any frictions and growth drivers that are present between time t and $t+1$, which may include adjustment frictions, changes in wages or productivity growth. The next two subsections thus focus on two key determinants of within-plant employment growth: productivity growth and plant-level hiring frictions.

Productivity growth and selection

What is the role of productivity growth in the Indonesian growth experience and how much productivity growth is explained by the better selection of plants? In this section, we show that aggregate productivity increased roughly 3.5-fold between 1975 to 2015. However, the underlying drivers of this productivity growth differ fundamentally across Indonesia's two main growth periods: during the period of rapid labor-intensive industrialization (1975-1997), all of the aggregate productivity gains are driven by the selection of more productive plants, while aggregate productivity gains during the period after the Asian Financial Crisis are almost entirely driven by within-plant productivity gains. We obtain these results by standard production function estimation (Ackerberg, Caves, and Frazer 2015; Demirer 2020) and then separately identifying the productivity improvements that come from the better selection of plants versus within-plant productivity improvements.

Estimating productivity Following the literature, we estimate a standard value-added Cobb-Douglas production function in capital k and efficiency units of labor h:

$$y_{it} = x_{it} h_{it}^{\theta_{jt}} k_{it}^{\alpha_{jt}} \tag{2.2}$$

with $\theta_{jt} + \alpha_{jt} \leq 1$ giving the output elasticities of labor and capital in sector j at time t and x_{it} capturing plant-level productivity. As a baseline and in line with our subsequent dynamic model, we start with common output elasticities across manufacturing industries, but discuss further industry variation below. We estimate the labor and capital output elasticities separately for each year allowing for both inputs to be potentially fully dynamically chosen, which means that at this point we can remain agnostic about the frictions that determine plant input choices and changes in these frictions over time. Specifically, we draw on the control function approach in Demirer (2020), which is close in spirit to the standard control function value-added production function estimation based on Ackerberg, Caves, and Frazer (2015), but does not require intermediate inputs (also see: Gandhi, Navarro, and Rivers 2017). We provide a discussion and an identification proof adapted to our setting in Appendix 2.A.5, but the intuition of identification is as follows: exploiting the assumption that productivity follows a first-order Markov process, conditional on previous input choices and output, the ranking of current inputs identifies the ranking of productivity innovations, which can be used to construct a control function for (unobserved) productivity in the output regression.

transition matrix that can be constructed between 1975 and 2015. An important caveat that we cannot address is that due to the discretization, the exercise would also give an increase in the plant size if all plants were simply growing at a constant rate. The red lines in Figure 2.6 report changes with a counterfactual transition matrix enforcing average employment growth rates between 1975 to 1976 across all plants. While observed within-plant employment growth is far from constant, the alternative exercise still shows that it is easy to overestimate the growth in the right tail.

Results and more details on the estimation are shown in Appendix 2.A.5. Before we discuss the results, it is important to highlight that we estimate the production function on available revenue data. This comes with the standard limitations that we only identify revenue elasticities and revenue-based productivity and cannot distinguish between productivity and demand nor identify changes in markups.[9]

We find no systematic changes in estimated output elasticities over time and very standard values for the output elasticity of labor close to 2/3. Importantly, the estimated output elasticity of labor is substantially larger than the median plant-level labor share (≈ 0.45) and the aggregate labor share in manufacturing (≈ 0.25). In a frictionless model with Cobb-Douglas production, plants would equalize cost shares and revenue elasticities. In the next section, we provide evidence on frictions for labor choices that could rationalize this difference. Our subsequent model then quantitatively accounts for these large differences. We find smaller estimates for the capital output elasticity than generally found in the literature, which – as we discuss in Appendix 2.A.5 – is likely due to attenuation bias from measurement error in observed capital. A lower capital output elasticity means that observed (mismeasured) variation in capital has smaller effects on output and we further show that mismeasurement of the capital elasticity is not biasing our estimates for labor.

Selection versus plant-level productivity growth Next, we quantify how much of the productivity improvements across plants are driven by the selection of more productive plants versus within-plant productivity growth. For this, we assume that plant-level productivity is the product of a common, potentially endogenous, aggregate technology component z_t that improves the productivity of all plants and an idiosyncratic productivity shock s_{it}: $x_{it} \equiv z_t s_{it}$. This setup allows us to separate shared technology growth in z_t, selection on idiosyncratic productivity s_{it} and within-plant growth in s_{it}, and nests the productivity side of many exogenous as well as endogenous growth models in the literature.[10] In this setup, we provide a novel non-parametric identification approach that separates the path of z_t from s_{it} (up to a normalization of z_0).

To understand why separate identification of selection and technology growth is difficult in

[9]Using separate information on prices and quantities, the previous literature has highlighted the important role of demand for driving firm growth (Hottman, Redding, and Weinstein 2016; Foster, Haltiwanger, and Syverson 2016; Eslava and Haltiwanger 2020) and an important part of what we subsequently call "productivity" likely captures demand. We return to this issue when discussing model counterfactuals where the distinction between demand and productivity is key. Also, revenue-based productivity measures may be preferred in the Indonesian context where large changes in product quality bias quantity-based productivity estimates (see Atkin, Khandelwal, and Osman (2019) for the argument and Hill (2000) for a discussion of strong quality improvements in Indonesian manufacturing). Relatedly, we also do not identify changes in mark-ups, which usually requires separate information on prices and quantities (Bond et al. 2021). Studying changes in markups and its drivers over the course of development is an exciting direction of future research, but beyond the scope of this paper.

[10]Specifically, the setup nests standard neoclassical growth models that feature exogenous aggregate productivity growth and firm selection (e.g. Luttmer 2007; Clementi and Palazzo 2016) as well as endogenous growth models that feature a common endogenous growth component such as Romer (1990) or models where Gibrat's law holds and productivity growth is independent of firm size (Klette and Kortum 2004; Atkeson and Burstein 2010; Restuccia and Bento 2015; Peters 2020).

the first place, let us start by looking at log changes in average productivity over time:

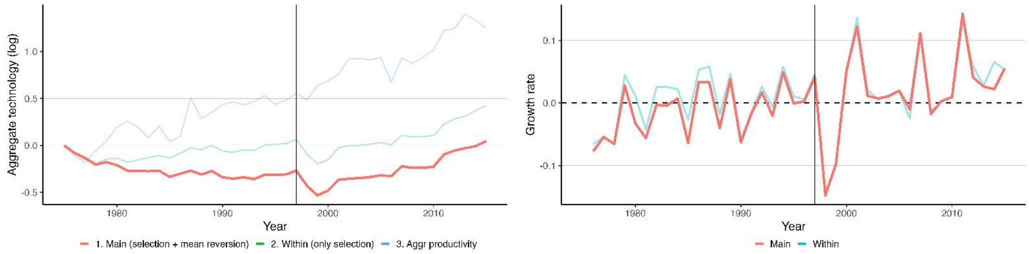
$$\frac{1}{N_t}\sum_{i\in\mathcal{N}_t}\tilde{x}_{it}-\frac{1}{N_{t-1}}\sum_{i\in\mathcal{N}_{t-1}}\tilde{x}_{it-1}=\underbrace{\tilde{z}_t-\tilde{z}_{t-1}}_{\Delta log(z)}+\underbrace{\frac{1}{N^S_{t,t-1}}\sum_{i\in\mathcal{N}^S_{t,t-1}}\Delta\tilde{s}_{it}}_{\text{Survivor }\Delta log(s)}+\underbrace{\frac{1}{N^E_t}\sum_{i\in\mathcal{N}^E_t}\tilde{s}_{it}}_{\text{Entry }\overline{log(s)}}-\underbrace{\frac{1}{N^X_t}\sum_{i\in\mathcal{N}^X_t}\tilde{s}_{it-1}}_{\text{Exit }\overline{log(s)}}$$

where $log(x)\equiv\tilde{x}$ denotes variables in logs. Changes in average productivity $\tilde{x}_{it}$ only identify growth in technology z under the special case that average changes in productivity $\tilde{s}$ among survivors as well as changes in average productivity $\tilde{s}$ between exiting and entering plants exactly cancel out. These terms capture two important selection biases. The entry and exit terms capture a "static" compositional selection bias that leads to overestimates of aggregate technology changes as long as less productive plants are more likely to exit and entering plants are positively selected on productivity. Focusing instead on within-plant changes in productivity deals with the "static" selection bias, but still leaves a term capturing average changes in idiosyncratic productivity among surviving plants. We call this term the "dynamic" selection bias, which generally biases estimated aggregate technology changes downwards if productivity s_{it} is persistent. Intuitively, if surviving plants are selected based on good histories of productivity realizations s_{it}, then they are more likely to mean revert in the future.

Non-parametric identification in our setup means that we make no functional form assumptions on the arbitrarily time-varying path of aggregate technology z_t, the productivity shock process s_{it}, nor on the plant entry and exit processes that drive endogenous selection. For expositional purposes, we provide an idea of the identification and estimation approach and relegate the precise technical assumptions, a detailed identification proof and further estimation details to Appendix 2.A.6. Identification of changes in z_t relies on two sets of assumptions. The first assumption ensures that the productivity shock process s_{it} has a stationary distribution. Technically, we assume that s_{it} follows the same underlying general first-order, ergodic Markov process across plants and time, allowing for flexible forms of error dependence. The stationary distribution is useful because if we could reweight changes in observed productivity x_{it} among the – potentially highly selected – set of surviving plants based on the stationary distribution of s_{it}, the "dynamic" selection bias exactly cancels out. That is, $\mathbb{E}_i\Delta\tilde{s}_{it}$ is exactly equal to zero at the stationary distribution of s. The second set of assumptions ensures that such a stationary distribution can always be constructed using the observed data, restricting the degree of selection at exit. Specifically, we require that (1) plants' exit decisions are not based on future productivity shock realizations, and (2) there is no sharp productivity cutoff at which all plants would exit, so that there is always common support that allows an appropriate reweighting of the distribution. The latter can be empirically tested and – as shown in Figure A.8 Panel B – finds strong support in the Indonesian data.

With the assumptions in hand, the only remaining difficulty is how to construct the weights of the stationary distribution and solve for the time path of changes in z_t. Here, we first solve "forward" for the stationary distribution by starting with equal weights over the initial distribution. Whenever a plant selectively exits, we pass on their weight to plants with similar productivity who did not exit using a standard Kernel estimator, creating a synthetic panel among surviving plants that is "representative" of the underlying process of s. For time growing large – no matter how selected the initial distribution is – one can thereby identify appropriate weights over the selected set of producing plants that recovers the stationary distribution

Figure 2.7: Aggregate manufacturing technology estimates



Notes: Panel A: Aggregate technology estimates, showing the main estimator (explained in the data) as well as the within estimator (that only controls for selection, but not mean reversion) and aggregate productivity (measured as value-added-weighted average productivity). Panel B: Corresponding growth rates in aggregate technology. The main and within estimators both use (weighted) median changes in plant-level productivity. Further details in the text.

of s (up to a common scaler z). We then move "backwards" from the last period T to identify the path of z_t: initially normalizing z_T, we start with a guess over z_{T-1}, compute the weighted changes in productivity between $T-1$ and T based on the stationary distribution of s in $T-1$ and solve for the implied z_T. This implies finding a root in z_{T-1}. Iterating on this procedure until z_0 recovers the entire path.

Figure 2.7 shows the estimated path of technology using the full sample and baseline productivity estimates. Over the entire 40-year period, technology improves little, being less than 5% higher in 2015 than in 1975. However, this masks important changes over time. Specifically, technology actually declined strongly throughout the 1970s, remained almost constant throughout the 1980s and 1990s and then saw rapid growth of roughly 4% per year since the year 2000. At the same time, aggregate productivity – measured as the value-added-weighted average productivity across plants – increased roughly 3.5-fold by 2015 and increased by 75% by the time of the Asian Financial Crisis in 1997. Together, this means that the sources of productivity growth fundamentally differed over the two Indonesian growth periods: plant selection rather than shared technology growth drove more than all of the productivity gains during Indonesia's period of rapid labor-intensive industrialization (1975-1997), while the pattern reversed after the year 2000, with common technology growth explaining more than 90% of the aggregate productivity gains. These estimates – especially after the Asian Financial Crisis – are driven by within-plant productivity growth of surviving plants. Pre-1997, true technology growth is lower than the unweighted within-plant productivity growth because of positive mean reversion that can be explained by the presence of many young and low productive plants that mean revert upwards in their productivity shocks. Post-1997, positive and negative mean reversion roughly balance out, explaining the similar growth paths of the main and within estimators. In Appendix 2.A.6 we further discuss what could be driving these large changes in the role of aggregate technology over time.

The allocation of resources did not improve over time

At last, we look at measures of the misallocation of resources over time. We start with an accounting-based decomposition of growth in Indonesian manufacturing. Using the previously assumed production structure and separation of plant-level TFP into an aggregate technology and idiosyncratic productivity component, we can write growth in manufacturing output as:[11]

$$\Delta\ ln\ Y_t \equiv \underbrace{\Delta\ ln \sum_i f(h_{it}, k_{it})}_{\text{input growth}} + \underbrace{\Delta\ ln\ z_t}_{\text{aggr techn.}} + \Delta\ ln \underbrace{\left[\bar{s}_t + N_t cov\left(s_{it}, \frac{f(h_{it}, k_{it})}{\sum_i f(h_{it}, k_{it})}\right)\right]}_{\text{selection + reallocation effect}} \quad (2.3)$$

where N_t tracks the number of active plants. Total output is the combination of the state of factor accumulation and aggregate TFP. Aggregate TFP, in turn, can be further decomposed into aggregate technology and – following Olley and Pakes (1996) – a combination of average productivity and a covariance term that captures whether resources in the economy are allocated towards the most productive plants. Since the covariance is affected by common trends in its inputs such as changes in the sample size, Figure 2.8) plots the correlation of plant-level productivity and resource shares, which is robust to common trends and simply the normalized covariance. Panel A and B show this measure of the allocation of resources for the full sample of plants and only for surviving plants that already operated in 1975, normalizing each by the first year. Each panel additionally plots the cross-sectional correlation within industries.[12] Across all plants, the allocation of resources actually deteriorated over time, while it remained stable for survivors, indicating that the deterioration is driven by the entry of small plants. As further documented in Appendix 2.A.7, this result also holds within a balanced panel and separately within each cohort of entering plants between 1970 to 1999.

In Appendix 2.A.7, we also report changes in the dispersion of marginal revenue products of capital and labor over time, which maps to changes in misallocation in the literature building on Hsieh and Klenow (2009).[13] Again, we find that, if anything, the dispersion of measured marginal revenue products (even within 5-digit industries) tends to increase over time. Thus, based on our estimates and in contrast to Buera and Shin (2013), we find little evidence for an undoing of misallocation being a feature of growth in Indonesia.

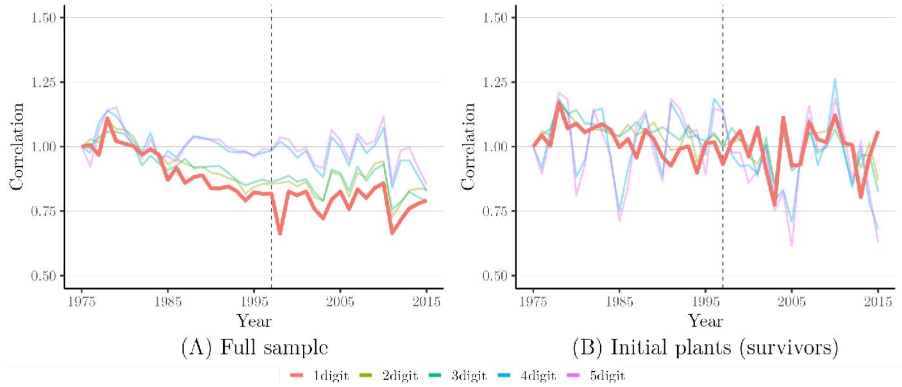
The misallocation dynamics shown in Figure 2.8) are a function of changes in plants' productivities and how inputs reallocate across plants. To shed light on these dynamics at the micro-level, we are interested in how plants' inputs respond to changes in productivity. Figure 2.9 shows how plants' labor and capital shares evolve as a response to a positive and permanent productivity shock of at least 20% – roughly equal to the 75th percentile of within-plant productivity changes. We use a standard staggered differences-in-differences design (Callaway and Sant'Anna 2021) that is particularly suited here, because it captures plant-level dynamic responses while controlling for time and plant fixed effects that ensure that results are neither driven by aggregate shocks in specific years nor by fixed differences across plants such as differences across industries or even plant-specific production functions. Given selective plant entry and exit and resulting composition biases from estimating treatment effects on un-

[11]The proof can be found in Appendix 2.B.4.

[12]Specifically, we proceed similar as in Gopinath et al. (2017): we first estimate the correlation across plants in a given industry and year and then construct the weighted average correlation across industries using the industry's average share in manufacturing value added as an industry-specific time-invariant weight. Using the same weights when aggregating across industries ensures that within-industry estimates reflect purely variation within industries over time.

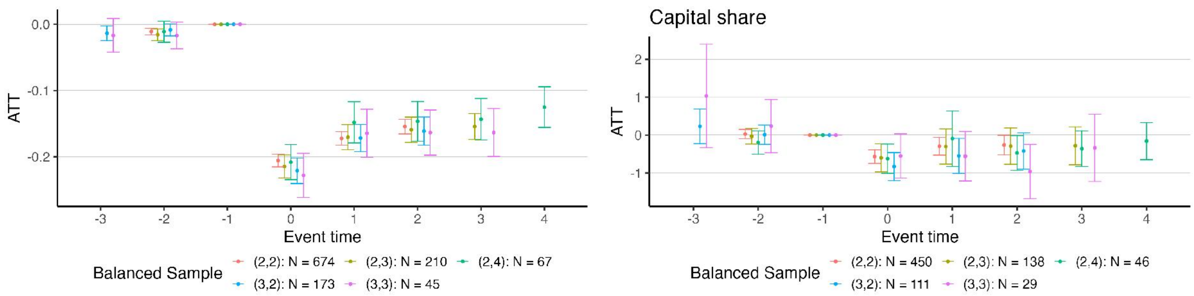
[13]There are important limitations to interpreting changes in the covariance between measured revenue productivity and input shares as changes in misallocation. For example, in Hsieh and Klenow (2009), an efficient allocation implies zero correlation between TFPR and the input share. We note that revenue productivity in our model is not equal to TFPR in Hsieh and Klenow (2009), but instead much more closely correlated with their measure of TFPQ.

Figure 2.8: Evolution of cross-sectional correlation of plant productivity and input share



Notes: Input shares are computed based on Cobb-Douglas aggregator. Within-industry results based on first estimating the correlation across plants in a given industry and year and then constructing the weighted average correlation across industries using the industry's average share in manufacturing value added as a time-invariant weight. All series are normalized by the first year.

Figure 2.9: Event study: Input share responses to a permanent productivity shock



Notes: Treatment defined by permanent productivity shock of 20 percent change that does not revert back, neither before nor after treatment. Event study following Callaway & Sant'Anna (2021) estimated on balanced panel of treated plants and with all non-treated plants as control, event time zero gives first period of treatment, effects at -1 normalized to zero by assumption. Sample for capital share is restricted to post 1990 due to data availability. Details in the text.

balanced panels, Figure 2.9 reports estimated dynamic treatment effects for different balanced panels.[14]

Plants' labor shares drop by more than 20 percentage points in the first year of treatment and recover slowly over time as plants respond by increasing hiring. In Appendix 2.A.8, we further show that the slow labor share adjustments are indeed driven by slow hiring rather than wage increases. In a model with the above production side but where plants can adjust employment without constraints, there should be no response in the labor share, indicating frictions in adjusting labor. Quantitatively, these frictions are large: For example, the average "treatment" shock is roughly 40%, so that a plant with an initial labor share of 0.7 would not adjust labor at all in the first year of a large productivity increase. In contrast, estimated dynamic responses

[14]The treatment effects are estimated with all non-treated plants as control. Event time zero gives the first period of treatment, while treatment effects in event time -1 are normalized to zero by assumption. Apart from the 20% cutoff, the treatment definition also ensures we identify a permanent productivity shock by ruling out shocks followed and preceded by productivity changes of more than 10% per year and ruling out an accumulation of shocks over multiple periods that undo the shock at period 0. E.g. if we look at an event window from -3 to 2, then productivity at -3 cannot be more than 10% apart from the pre-level at -1 and productivity at 2 cannot be more than 10% apart from productivity at event 0. This does not restrict input shares. The rapid observed declines in the sample sizes are due to the high observed volatility of productivity, making it difficult to identify permanent shocks in the data.

for capital are much noisier and we cannot reject that capital shares immediately recover after the first year. For both inputs, we find no evidence for pre-trends, pointing away from an anticipation of productivity shocks. Taken together, volatile plant-level productivity and slow labor adjustments thus offer an explanation as to why the allocation of resources did not improve over time.

2.3 Structural model

While the four empirical facts document important changes in the economy and distribution of manufacturing plants over time (Fact 1), entry and exit dynamics that shape the selection of plants (Facts 2, 3 & 4), and the importance of slow plant-level adjustment processes (Facts 2 & 4), the empirical evidence is not sufficient to quantify the drivers of the Indonesian growth miracle. In particular, it does not allow us to quantify the aggregate effects of policy changes separately from transition growth. For this, we now build a model of plant dynamics and growth in the tradition of Hopenhayn (1992). In the model, plants face idiosyncratic risk in their productivity and choose capital and labor inputs subject to labor adjustment costs and a simple financing constraint that rationalize slow plant-level adjustments (Facts 2 & 4). Plants face fixed costs that drive endogenous entry and exit, which in turn drives aggregate selection dynamics (Fact 3).

The model features a time-varying growth path which is driven by three endogenous forces: changes in the input distribution, changes in the productivity distribution due to a combination of exogenous technology growth and plant selection, and changes in (mis)allocation as given by their joint distribution. All three forces are driven by the race between transition growth and by changes in model fundamentals that induce new transition growth. Changes in model fundamentals include changes in labor supply, potential entrants, technology, adjustment frictions and taxes. We further embed the model of plant heterogeneity into a two-sector general equilibrium model to capture changes in the rest of the economy and the endogenous reallocation of labor across sectors over time. The potential of transition growth at any point in time is given by the current distribution of plants encoding the history of the economy and future growth potential as given by current model fundamentals and expectations over the future. We follow the growth literature in treating the growth path as deterministic and by assuming that agents have perfect foresight over aggregate changes in the economy.

2.3.1 Model Setup

The model economy is set in discrete time indexed by $t = 1, 2, ...$. We assume that Indonesia is a small open economy vis-a-vis the rest of the world and has access to world capital markets at interest rate r^*. There are two sectors of production: Manufacturing (M) and the rest-of-the-economy (R). Both sectors produce the same homogeneous, perfectly substitutable good, which serves as numeraire. Manufacturing features heterogeneous plants whose endogenous mass and distribution are time-varying, while we model the rest-of-the-economy as a simple representative firm whose exogenous technology and endogenous labor demand change over time. Labor is inelastically supplied by households which choose in which sector to work. Labor markets in both sectors are fully competitive. There is a government that levies a valued added and a corporate income tax, the two main corporate tax instruments in Indonesia.

We assume the government levies these taxes and redistributes revenue back to households. Similarly, we assume that all plant profits are simply transferred back to households.

Manufacturing

The manufacturing sector is composed of risk-neutral plants that are heterogeneous in their productivity s_{it}. Each period, plants choose capital k and labor h on spot markets to produce output, while facing idiosyncratic risk over their future productivity s_{it} and time-varying changes in the economy. We denote plants' payoff-relevant aggregate state of the economy, including perfect anticipation of future changes in aggregates by Ω_t and make its components more explicit below. A plant's output y_{it} and taxable profits π_{it} at time t are given by:

$$y_{it}(s_{it}, z_t, h_{i,t}, k_{i,t}) = z_t s_{it} h_{it}^{\theta} k_{it}^{\alpha} \tag{2.4}$$

$$\pi_t(s_{it}, h_{i,t}, k_{i,t}; z_t, w_t) = (1-\tau_t^C)\big((1-\tau_t^{VAT})y_{it}(s_{it}, z_t, h_{i,t}, k_{i,t}) - w_t h_{i,t} - R_t k_{i,t}\big) \tag{2.5}$$

where the production function is as in Section 2.2.2, τ_t^{VAT} gives the Indonesian value-added tax and τ_t^C the corporate income tax that is levied on taxable profits. Given the frictions in this economy, both tax instruments generally distort input choices. R_t gives plants' capital borrowing rate, which is equal to the deposit rate r_t plus depreciation δ assuming competitive rental markets. Idiosyncratic risk s_{it} follows a Markov process of order one. We further assume that s_{it} is exogenous and independent of aggregate technology z_t. For z_t, we leave the path unrestricted, but assume it is exogenous.

What drives slow plant adjustments? We assume that plants face labor adjustment costs. These capture, for example, managerial time constraints that arise from the time it takes to hire, fire and reorganize production tasks, a key constraint for plant growth in developing countries (Bloom et al. 2013, 2020).[15] Following the literature, we model them as follows (e.g. Cooper, Gong, and Yan 2018; Cooper, Haltiwanger, and Willis 2015):

$$AC(h_{i,t-1}, h_{i,t}) = \begin{cases} F^+ + c_0^+(h_{i,t} - h_{i,t-1}) + \frac{c_1^+}{2}\left(\frac{h_{i,t}-h_{i,t-1}}{h_{i,t-1}}\right)^2 h_{i,t-1} & \text{if } h_t > h_{t-1} \\ 0 & \text{if } h_t = h_{t-1} \\ F^- + c_0^-(h_{i,t-1} - h_{i,t}) + \frac{c_1^-}{2}\left(\frac{h_{i,t}-h_{i,t-1}}{h_{i,t-1}}\right)^2 h_{i,t-1} & \text{if } h_t < h_{t-1} \end{cases} \tag{2.6}$$

where F are fixed adjustment costs that capture overhead in dealing with hiring (F^+) or firing (F^-) and c_0 captures per worker hiring and firing costs. Importantly, there are convex adjustment costs whose importance is captured by c_1 and which capture costs of growing (c_1^+) or shrinking (c_1^-) plants quickly. Convex adjustment costs are key to explain the slow growth of plants over time and are a key determinant of the speed of transition growth. We allow all costs to be asymmetric to accommodate that firing and hiring is often regulated differently. At last, we index all adjustment costs in terms of wages since wage indexation provides a simple way to let costs grow with the economy. Besides this indexation, in the baseline model, we

[15] In Appendix 2.B.1, we provide a simple microfoundation in terms of the costs of scarce managerial time to show how organizational changes induce convex costs. An alternative interpretation of convex adjustment costs is given in labor search models where they are rationalized via convex (reduced-form) hiring or vacancy posting costs (e.g. Bilal et al. 2022; Coşar, Guner, and Tybout 2016)). The key difference is that adjustment costs in search models become partly functions of equilibrium outcomes such as market tightness. We abstract from this general equilibrium mechanism here given that the primary focus of the paper is on longer run growth dynamics and not business cycle variation in unemployment and market tightness.

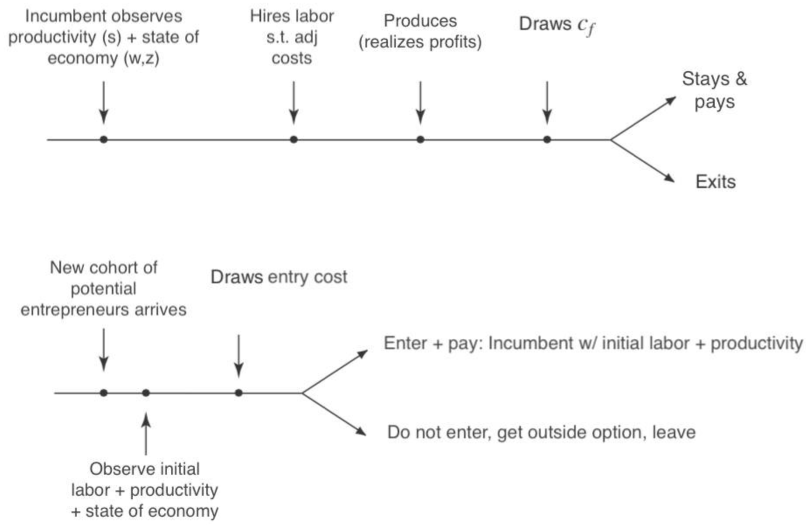


Figure 2.10: Timing in time t for manufacturing firms.

assume that adjustment cost parameters are fixed over time, but the model and estimation can accommodate time-varying parameters and we return to this point in later sections.

On top of adjustment costs, we assume that plants face a standard financing constraint that captures underdeveloped financial markets, in line with a large Macro Finance literature that emphasizes their importance (e.g. Buera and Shin 2013; Midrigan and Xu 2014; Moll 2014). Specifically, we assume the following working capital constraint:

$$w_t h_{it} \leq \kappa_t (1 - \tau_t^{VAT}) y_{it} \quad \text{(Working capital constraint)} \tag{2.7}$$

As we show in Appendix 2.B.2, the constraint can be microfounded as a simple limited enforcement problem whereby plants may want to run deficits in order to grow their workforce, but cannot commit to repaying. Limited contract enforcement due to a weak judicial system then leads to the borrowing constraint in equilibrium. κ_t then captures the probability that the judicial system will enforce the contract, with higher κ_t mapping to stronger institutions and a less binding constraint. The reason that plants may want to run deficits to grow their workforce is that in the presence of convex adjustment costs it is costly to build up a plant's workforce quickly, so that even less productive plants may want to run deficits to build up their workforce hoping for good future productivity realizations.

The timing of manufacturing production is summarized in Figure 2.10. At the beginning of a period, incumbents observe their current productivity and make production decisions. Plants' payoff-relevant aggregate state is given by: $\Omega_t \equiv \{z_t, w_t, r^*\}_t^\infty$. After production takes place, incumbent plants incur a fixed cost of production $c_{i,t}^F$, upon which plants decide whether they want to continue producing (and pay $c_{i,t}^F$) or permanently exit (as in Clementi and Palazzo 2016). The fixed cost is drawn from a distribution G, which we assume to be Gumbel with scale and variance parameters (μ_t^X, σ_t^X). A larger variance in the fixed costs rationalizes more overlap in the labor and productivity distributions of surviving and exiting plants. The exit decision of the plant depends on the plant's expected future value, the cost shock as well as the costs of closing down the plant (as in Hopenhayn and Rogerson 1993):

$$\max\left\{\beta\mathbb{E}[V^M(s_{i,t+1},h_{i,t};\Omega_{t+1})|s_{i,t},h_{i,t};\Omega_t]-c^F_{i,t},-C_E(h_{i,t})\right\} \tag{2.8}$$

where V^M gives the continuation value of an incumbent plant and $C_E(h_{i,t})$ the costs of closing down the plant. The above maximization problem implicitly defines plant i's survival probability that the operating cost draw $c^F_{i,t}$ is lower than its future expected continuation value: $\lambda(s_{i,t},h_{i,t};\Omega_t)\equiv\mathbb{P}\big(x\geq c^F_{i,t}\big)=G\big(x\big)$ where $x\equiv\beta\mathbb{E}[V^M(s_{i,t+1},h_{i,t};\Omega_{t+1})|s_{i,t},h_{i,t};\Omega_t]+C_E(h_{i,t})$.

The ex-ante value of an incumbent manufacturing plant can be written in recursive form according to:

$$\begin{aligned}V^M(s_{i,t},h_{i,t-1};\Omega_t)=\max_{h_{i,t}\leq\bar{h},k_{i,t}}&\Bigg\{\pi(s_{it},h_{i,t},k_{i,t};z_t,w_t)-w_tAC_t(h_{i,t-1},h_{i,t})+\lambda(s_{i,t},h_{i,t};\Omega_t)\Big\{\\&-\mathbb{E}_c[c^F_{it}|\text{stay}(s_{i,t},h_{i,t};\Omega_t)]+\beta\mathbb{E}\big[V^M(s_{i,t+1},h_{i,t},\Omega_{t+1})|s_{i,t},h_{i,t},\Omega_{t+1}\big]\Big\}\Bigg\}\end{aligned} \tag{2.9}$$

Plants have a common discount factor $\beta=1/(1+r^*)$, which is pinned down by the world interest rate. The presence of adjustment costs and financial constraints in combination with productivity dynamics makes this a dynamic problem since plants take into account that contemporaneous changes in their inputs influence adjustment and financing costs in the future.

Next, we consider endogenous plant entry. As visualized in Figure 2.10, each period there is a cohort of potential entrants (PE) of measure $|PE|_t$. Each potential entrant draws a random entry cost c^E_{it} from a distribution P, which they need to pay in case they start producing. Again, we assume that entry costs follow a Gumbel distribution with scale and variance parameters (μ^E_t,σ^E_t). Potential entrants differ in their idiosyncratic productivity s_{it} and their initial labor $h_{i,t}$, which they know when making the entry decision for producing in period t. The initial heterogeneous level of labor and productivity is key to capture that there is plant entry of many small as well as some very large plants that matter in the aggregate. It also accounts for the fact that we only observe and model plants with 20 or more workers. The exogenous distribution of potential entrants is given by $PE_t(h_t,s_t)$, which is time-varying due to exogenous reasons such as demographic changes that have been shown to be key for explaining variation in firm creation over time (Bernstein et al. 2022; Karahan, Pugsley, and Şahin 2019; Liang, Wang, and Lazear 2018). Potential entrant i with entry cost shock c^E_{it} enters if its net value is positive:

$$V_{PE}(s_{i,t},h_{i,t};\Omega_t)=\max\left\{V^M(s_{i,t},h_{i,t};\Omega_t)-c^E_{it},0\right\} \tag{2.10}$$

where we have normalized the outside option to zero. Similar to exit, this gives the following mapping $\mathbb{P}\big(V^M(s_{i,t},h_{i,t};\Omega_t)\geq c^E_{it}\big)=P\big(V^M(s_{i,t},h_{i,t};\Omega_t)\big)$. Note that in this specification the initial mass and distribution of entrants is endogenous, but entrants only start making input choices the period after they entered. A time-varying distribution of potential entrants also allows us to deal with plant entry jumps in years of the economic census as we further discuss in the estimation section. We denote the endogenous mass of entry for each state (h_t,s_t) in period t by $\mu(h_t,s_t)$, which is a function of Ω_t. Similarly, we define by $m(h_t,s_t)$ the endogenous mass of producing plants for each state (h_t,s_t) in period t. With slight abuse of notation, denote by M_t the set of producing plants at time t.

Rest-of-the-economy

We model the rest-of-the-economy parsimoniously as a representative firm with a decreasing returns to scale (DRS) production function:

$$Y_t^R = A_t \left(H_t^R\right)^{\theta_R} \quad \text{with} \quad \theta_R \in (0,1) \tag{2.11}$$

where A_t is time-varying TFP, H_t^R gives labor employed and θ_R gives the output elasticity in the rest-of-the-economy. Decreasing returns to scale ensure that economy-wide wages can be affected by changes in manufacturing. The rest-of-the-economy sector takes as given productivity A_t and the wage rate w_t and chooses optimal labor demand maximizing per period profits: $\pi_t^R(A_t, w_t, \tau_t) = Y_t^R(A_t) - (1+\tau_t^R)w_t H_t^R$, subject to labor demand wedges τ_t^R. Labor demand is then given by:

$$H_t^{R*} = \left(\frac{\theta_R A_t}{(1+\tau_t^R)w_t}\right)^{\frac{1}{1-\theta_R}} \tag{2.12}$$

Labor demand wedges in the rest-of-the-economy are a simple way to capture observed variation in the labor intensity of output and one can think of them as changes in labor frictions. A_t captures changes in technology of the rest-of-the-economy. We allow both A_t and τ_t^R to change over time, but treat them as deterministic and exogenous paths.

Households

There is a continuum of households j that are characterized by their exogenous household-specific efficiency units of labor h_{jt} and whose exogenous mass at time t is denoted by L_t. Households supply labor inelastically so that the aggregate labor supply is given by $H_t = \int_j h_{jt} dj$. Changes in L_t and $\frac{H_t}{L_t}$ capture changes in the working population and education per worker respectively. We abstract from consumption-savings decisions by assuming that households are hand-to-mouth, simply consuming their labor income y_{jt} net of transfers from the government and plant profits T_{jt}: $c_{jt} = y_{jt} + T_{jt}$.[16] Households allocate their labor supply across both sectors based on maximizing labor income: $y_{jt} = \max\{h_{jt}w_t^M, h_{jt}w_t^R\}$.

Equilibrium

We assume that the observed growth path in the data is characterized by a path of per-period perfect foresight *Recursive Competitive Equilibria*.

Definition 5 (Model fundamentals.)**.** *Model fundamentals at time t capture all exogenous model parameters, processes and distributions as given by:*

$\Theta_t^F = \{\theta, \alpha, \delta, F^-, c_0^-, c_1^-, F^+, c_0^+, c_1^+, \{A_t, \tau_t^R, H_t, PE_t, z_t, \tau_t^C, \tau_t^{VAT}, \kappa_t, \mu_t^X, \sigma_t^X, \mu_t^E, \sigma_t^E\}_t^\infty\}$.

We further denote by $\bar{\Theta}_t^F$ the modified set of model fundamentals where all fundamentals are fixed to

[16]We only make this assumption to fix ideas. Given the small open economy setup, the domestic supply of capital is inelastic to changes in domestic savings behavior such that the production side – which is the focus of this paper – would look exactly the same if heterogeneous households would instead solve a savings and consumption choice. Given inelastic labor supply, the model is also isomorphic to one with uniform labor income taxes in both sectors.

their value at time t forever. At last, denote by $\Theta_t^F \setminus \{x = \bar{x}\}$ the modified set of model fundamentals where only model fundamental x is changed to $\bar{x}$.

Definition 6 (Initial distribution.)**.** *The distribution of surviving plants from period $t-1$ gives the initial distribution at time t and is denoted by S_t.*

A path of perfect foresight *Recursive Competitive Equilibria* starting at time t is then given by model fundamentals Θ_t^F, an initial distribution S_t, and endogenous sequences of prices $\{w_t, r^*\}_t^\infty$, corresponding quantities and distribution of producing plants $\{m_t\}_t^\infty$ such that each period $\tau \in [t, \infty)$:

1. The rest-of-the-economy sector statically chooses optimal labor demand maximizing profits taking as given productivity A_τ, the wage w_τ and labor demand wedges τ_τ^R.
2. Manufacturing plants choose optimal labor and capital demand.
3. Potential entrants optimally make entry and incumbents optimally make exit decisions.
4. Households inelastically supply total labor H_τ and optimally allocate labor across sectors to maximize labor income.
5. The aggregate wage w_τ adjusts to ensure that the labor market clears: $H_\tau = H_\tau^R(w_\tau, A_\tau, \tau_\tau^R) + \sum_{i \in M_\tau} h(s_{i\tau}, h_{i,\tau-1}; \Omega_\tau)$
6. The government runs a balanced budget by levying a value added and corporate income tax and redistributing revenue back to households.
7. The capital market clears every period such that international capital supply equals domestic capital demand: $\sum_{i \in M_\tau} k(s_{i\tau}, h_{i,\tau-1}; \Omega_\tau) = K_\tau^{INT}$.
8. The mass of active plants in τ and previous aggregate state $\Omega_{\tau-1}$ is equal to surviving plants from $\tau - 1$ plus endogenous new entrants:
$$\forall (s_\tau, h_\tau): \quad m(s_\tau, h_\tau; \Omega_\tau) = \sum_{s_{\tau-1}, h_{\tau-1}} \Big(\mathbb{1}_{h^* = h_\tau} \mathbb{P}[s_\tau | s_{\tau-1}] \lambda(s_{\tau-1}, h_{\tau-1}; \Omega_{\tau-1}) \times m(s_{\tau-1}, h_{\tau-1}; \Omega_{\tau-1}) \Big) + \mu(s_\tau, h_\tau) \tag{2.13}$$
9. The goods market clears each period such that total production is either consumed or exported: $Y_\tau = \sum_{i \in M_\tau} y_{i,\tau} + Y_\tau^R = C_\tau + NX_\tau$ where $NX_\tau = EXP_\tau - K_\tau^{INT}$ are net exports.[17]

The observed growth path features a combination of changes in model fundamentals that move the economy's *steady state* and transition growth as the economy is trying to catch up to this steady state. We now formalize these concepts.

Definition 7 (Balanced Growth Path (BGP) and Steady State (SS))**.** *Along a* BGP*, underlying technology in both sectors (A_t, z_t) and the endogenous wage grow at the same constant rate, while all remaining model fundamentals and the endogenous distribution of plants stays constant (see details in*

[17]Note that we have implicitly treated all entry costs, fixed costs and adjustment costs as shadow costs here, as they neither directly enter labor market clearing nor the aggregate resource constraint.

Appendix 2.B.3). A steady-state *is a BGP for which the growth rate is zero. Both* BGP *and* steady-state *are uniquely defined by* model fundamentals Θ^F *that admit a BGP/SS.*[18]

Definition 8 (Transition path)**.** *The unique perfect foresight* transition path *starting at t towards a BGP is defined by an initial distribution S_t and model fundamentals Θ_t^F (which admit a BGP), and gives a path of equilibrium wages over the transition.*[19]

In Section 4, we use the model to separately quantify the role of *transition growth* from changes in the *steady state*.

2.3.2 Estimation

The model captures a race between changes in model fundamentals and the distribution of plants (over the state space) trying to catch up to these changes. We now take this model to the data and show how to disentangle changes in the distribution from changes in model fundamentals. Estimation proceeds in three main steps, while an additional step is needed for model counterfactuals. The first step identifies equilibrium prices – only wages in our case – in the data. We take an equilibrium estimation approach (Hotz and Miller 1993; Bajari, Benkard, and Levin 2007; Caliendo, Dvorkin, and Parro 2019), which means that we treat our model as generating the equilibrium wage path we observe in the data. We can thus treat the path of equilibrium wages as fixed throughout the estimation and only need to solve for changes in the equilibrium wage path for counterfactuals. This greatly simplifies the estimation as it avoids solving for the equilibrium path of the model during the estimation. In the second step, we identify the distribution over the entire state space of the economy over time and use this to back out related model fundamentals such as the initial distribution. The third step then solves for remaining fundamentals that are related to the dynamic input and exit choices of plants drawing on observed choices of plants conditional on the state space. In this step, we also need to make an explicit assumption about the evolution of model fundamentals beyond the time frame of our data. To conduct model counterfactuals, we further back out model fundamentals that are not needed to solve the baseline economy, such as fundamentals of the rest of the economy.

With each estimation step, we enforce more model structure and assumptions, making parameter identification very transparent. An important benefit of our approach is that we can directly draw on the production function and aggregate technology estimates discussed in Section 2.2.2 whose identifying assumptions nest our model. Table 2.1 provides an overview of the estimation steps and all model fundamentals and estimates. We now discuss each step in more detail. Given the larger number of model fundamentals, we focus throughout on the most important parts and relegate details to Appendix 2.B.5.

[18] Uniqueness depends on the uniqueness of the individual policy functions and the unique mapping between policies and prices. We treat our numeric algorithm as formally defining the equilibrium refinement conditions sufficient for uniqueness.

[19] Uniqueness of the *transition path* can be proven via contraction mapping arguments over the path of price expectations. Again, our numeric algorithm for the transition gives a unique path for perfect foresight equilibria and we treat this as formally defining the equilibrium refinement conditions sufficient for uniqueness.

Table 2.1: Overview of parameter identification and estimation

Object	Description	Type	Identification idea	Value	Details
<u>Parameterization:</u>					
r^*	World interest rate	F	Risk-free rate	0.04	
δ	Depreciation rate	F	Standard	0.1	
τ_t^C	Corporate tax	F	Official rate	0.2	Section A.2.4
τ_t^{VAT}	VAT	F	Official rate	0.1	Section A.2.4
<u>Estimation:</u>					
Step 1:					
θ	Prod function	F	Control function	0.694	Section 2.3.2
α	Prod function	F	Control function	0.03	Section 2.3.2
w_t	Wage path	E	$\Delta_i(w_t h_{it})/l_{it}$	Fig. 12	Section 3.2.2
κ_t	Borrowing constraint	F	Max labor share	1.7	Section A.2.4
Step 2:					
z_t	Techn path	F	$\Delta_i$productivity$_{it}$	Fig. 8	Section 2.3.2
$\mathbb{P}(s'\|s)$	Transition matrix	F	Obs. transitions		Section 3.2.2
Init distrib		F	Obs. survivors		Section 3.2.2
E_t	Entrants	E	Obs. entry		Section 3.2.2
Step 3:					
Adj costs		F	Euler CCC	Table 2	Section 3.2.2
Cost ratio	Fixed cost	F	Euler CCC	Table 2	Section 3.2.2
Cost level		F	Match mass 2015		Section 3.2.2
For counterfactuals:					
A_t, θ_R, τ_t^R	Rest-of-Economy	F	First-order condition		Section A.2.4
Entry costs		F			Section A.2.4
PE_t	Potential Entrants	F	Entrants + entry proba		Section A.2.4

<u>*Details:*</u> Types are: F(undamental) and E(quilibrium object). The former stay fixed in counterfactuals, the latter change endogenously. If applicable, reported standard errors correct for multi-step estimation procedure and cluster at the plant-level by using block bootstrap across all estimation steps (This is currently still work in progress).

Step 1: Equilibrium wage estimation

In the first step, we estimate the path of equilibrium wages that clear labor markets in each period. While prices could in principle be directly observed in the data, based on our model, we only observe plants' wage bills ($w_t h_{it}$), which are a combination of the wage and the quantity of labor. We are only interested in changes over time and thus normalize the level of initial wages w_0 to unity. Ideally, we would like to capture changes in the wage by looking at wage changes for a worker whose efficiency units of labor remained constant. This would for example avoid any assumptions on how workers with different skills select across plants. In the absence of worker-level data that spans the entire time period, we instead draw on changes in within-plant per worker wages for similar job types, exploiting that the Indonesian data reports wages and the number of workers separately for production and non-production work. Our identification strategy for the wage allows for arbitrary sorting of workers with different skills not only across plants but also across different job types within plants, but restricts changes in the skill sorting within job types over time. Formally, we assume that plant i uses on average the same skills per worker within job types k: $h_{it}^k / l_{it}^k = \alpha_i^k \cdot \varepsilon_{it}^k$. ε_{it}^k allows job types within plants to vary in their skill intensity around α_i^j over time. With standard restrictions on ε_{it}^k, this ensures that wages are identified from:

$$\mathbb{E}_i \left[log(w_{t+1} h_{it+1}^k / l_{it+1}^k) - log(w_t h_{it}^k / l_{it}^k) \right] = log(w_{t+1}) - log(w_t)$$

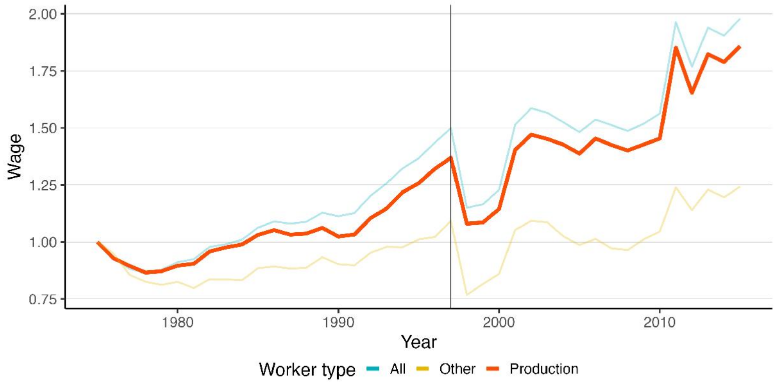
As our estimate of changes in wages, we use median within-plant-worker-type changes in wages, weighting observations by the average of total workers of type k between t and $t+1$, ensuring that wages are identified from median wage changes of workers (not plants). If anything, we think this estimator overestimates wage increases, because (1) any increase in within-worker human capital (e.g. on-the-job learning) will be attributed to increases in the wage, and (2) the estimates are for surviving plants, which might see more wage growth.

Figure 2.11 plots the estimated real wage series in the data. Our preferred estimator uses wage bills for production workers only, as production workers are relatively homogeneous and thus the identifying assumption is more likely to hold. For completeness, we also report estimated wage series using non-production workers and pooling all workers. Our estimates show that wages per efficiency unit of labor increased by more than 85% over the 40-year period with important variation over time. Given that the average wage bill per manufacturing worker increased roughly 4-fold in the data, these estimates imply that the average manufacturing worker in 2015 was about 2.2x more efficient than the average manufacturing worker in 1975. We find this a reasonable estimate given the large educational gains of Indonesian workers observed over this period. As external validation, Gathen (2021) also finds a similar wage increase from separate estimates on Indonesian worker data between 1998 and 2015.

Step 2: Mapping the entire distribution over the state space

In the second step, we map any plant in our data to the state space of our model: $(h_{it-1}, s_{it}, w_t, z_t)$. While our data captures a discrete number of plants, we treat this mass as continuous for the estimation. This allows us to identify changes in the entire distribution of plants over the state space over time, which is crucial to determine the potential for transition growth. Specifically, we use this mapping to identify two key model fundamentals – the productivity process and

Figure 2.11: Evolution of estimated real wage in Indonesian manufacturing



Notes: Based on within-task changes in the wage bill per worker by task (production and non-production). Data: Indonesian manufacuring census (1975-2015, 20+ workers).

the initial distribution – and an equilibrium object which changes in counterfactuals: the distribution of entrants.

To identify the state space, we draw on the wage estimates from Step 1, which together with a plant's wage bill ($w_t h_{it}$) identifies h_{it-1}. To obtain plant productivity and technology, we draw on the identification and estimation approach from Section 2.2.2, which hold under weak assumptions on exit and input choices. Having identified plant productivity s_{it} (up to a normalization of z_t), we estimate the dynamic process of s_{it} by discretizing s and then estimating the transition matrix $\mathbb{P}(s'|s)$ non-parametrically using (pooled) within-plant productivity changes in the data. The transition matrix is a fundamental of the economy and is identified based on within-plant changes in productivity conditional on previous productivity, and only requires that all productivity states are observed with positive probability at some point.[20]

Next, we exploit the state space mapping to identify the initial distribution of surviving plants over $(s_{i,t}, h_{i,t-1}; \mathbf{\Omega})$ in 1976. While our data starts in 1975, the first year for which we can identify $h_{i,t-1}$ is 1976. We treat this initial distribution as a model fundamental, implicitly assuming that in any counterfactual, initial survivors do not anticipate any changes to the baseline equilibrium paths of wages and technology prior to 1976. The main benefit of directly taking the initial distribution from the data is that we can remain agnostic about its origins and allow the data to reveal the initial degree of misallocation. A downside is that if the model does not capture all mechanisms of dispersion over the state space, the initial distribution may look more "misallocated" than it actually is; leading to overestimating model-implied transition dynamics.

At last, the state space mapping also allows us to identify time-varying entrant distributions $E(s_t, h_t; \mathbf{\Omega})$, which are equilibrium objects but related to the fundamental potential entrant distributions via: $PE_t(s_t, h_t; \mathbf{\Omega}) = E_t(s_t, h_t; \mathbf{\Omega}) / \mathbb{P}_E(s_t, h_t; \mathbf{\Omega})$, where $\mathbb{P}_E(.)$ gives the entry probability, which is a function of the model-implied value of entering as well as the parameters of the entry cost distribution. For the model estimation along the baseline equilibrium path, we treat these equilibrium entrant distributions as fixed. Our approach implies that the baseline

[20]While an imperfect ex-post test, we check ergodicity of the implied idiosyncratic productivity process s_{it} by verifying that all states in the discretized transition matrix can be eventually reached. In Appendix 2.B.5, we provide details on the discretization of productivity and labor, which we rely on for numerically solving the model and counterfactuals.

model exactly replicates observed plant entry. This is in contrast to plant-level exit and labor demand decisions, for which our estimation approach allows the model to fail.

Step 3: Estimating the dynamics of the model

Step 3 reveals the remaining parameters of the economy that are needed for the baseline model: fixed cost parameters that govern entry and exit decisions as well as adjustment cost parameters that govern how plants make dynamic labor choices. This step enforces more model structure, particularly on how plants make dynamic input, exit and entry choices, and how plants form expectations over the future. We separate this section into two parts with the first part only exploiting optimal plant-level choices across consecutive periods, while the second part also enforces long-run expectations.

Euler equation CCC estimation We identify most remaining parameters by exploiting observed exit and labor input choices conditional on the state space, drawing on standard conditional choice probability (CCP) and continuous conditional choice (CCC) Euler estimation techniques (Hotz and Miller 1993; Bajari, Benkard, and Levin 2007). Taking first-order conditions with respect to labor from the incumbent's value function above and directly plugging in the envelope condition, we obtain the following Euler equation:

$$
\begin{aligned}
0 = & \underbrace{\frac{\partial \pi(s_{i,t}, k_{it}, h_{i,t}, z_t)}{\partial h_{i,t}}}_{\text{Labor wedge}} - \underbrace{w_t \frac{\partial C_h(h_{i,t}, h_{i,t-1})}{\partial h_{i,t}}}_{\text{Current marginal adj costs}} \\
& + \underbrace{\frac{\partial \lambda(s_{i,t}, h_{i,t}, \Omega_t)}{\partial h_{i,t}} \left\{ -\tilde{g}(s_{i,t}, h_{i,t}, \Omega_t) + \beta \mathbb{E}\left[V(s_{i,t+1}, h_{i,t}, \Omega_{t+1}) | s_{i,t}, h_{i,t}, \Omega_t \right] \right\}}_{\text{Marginal benefit on survival}} \\
& + \underbrace{\lambda(s_{i,t}, h_{i,t}, \Omega_t) \left\{ -\frac{\partial \tilde{g}(s_{i,t}, h_{i,t}, \Omega_t)}{\partial h_{i,t}} + \beta \mathbb{E}\left[-w_{t+1} \frac{\partial C_h(h_{i,t+1}, h_{i,t})}{\partial h_{i,t}} | s_{i,t}, h_{i,t}, \Omega_t \right] \right\}}_{\text{Marginal benefits on future costs}}
\end{aligned}
\tag{2.14}
$$

where we have used $\tilde{g}(s_{i,t}, h_{i,t}, \Omega_t)$ to denote the expected fixed cost conditional on surviving to emphasize that it is a function of the state space.

The Euler equation, which holds for any plant that is optimally adjusting labor, says that plants should equalize today's marginal product of labor with the marginal costs of labor and current as well as future labor adjustments. Adjustment costs give a natural explanation for why there is a "wedge" between the static marginal product and the marginal costs of labor (Hsieh and Klenow 2009). For our estimation purposes, the important features of the Euler equation are that it holds along the transition, and that it gives a nonlinear equation in observable plant-level choices (exit and input choices) and parameters that govern survival probabilities as well as adjustment costs. Specifically, marginal adjustment costs are a function of adjustment cost parameters. The tricky terms are expected and marginal expected fixed costs, marginal survival probabilities and the expected future continuation value. As we show in Appendix 2.B.6, the Gumbel distribution for the fixed costs ensures that we can analytically invert all of these terms as functions of observed survival probabilities and parameters of the Gumbel distribution.

Appendix Table B.1 presents the non-linear least squares (NLS) estimation results and Appendix 2.B.6 gives the exact estimating equation and estimation details. The Euler equation flexibly identifies marginal adjustment costs. Intuitively, linear adjustment costs are identified from the observed labor wedge across plants and the probability of switching between shrinking and growing as determined by the volatility of the estimated productivity process. Convex costs instead scale with the labor growth and are thus identified from the variation in within-plant labor demand growth across periods, again conditioned by the observed volatility of the productivity process. We find sizable adjustment costs – especially convex costs on growing – that rationalize why even productive plants (with a high labor wedge) conditional on previous plant employment do not grow faster. Quantitatively, our estimates imply that growing a plant's workforce by 20% within a year – a growth rate slightly above the 75th percentile – leads to adjustment costs that are about 75% of the previous wage bill. Informed by faster observed shrinking conditional on productivity, convex costs on shrinking are estimated to be less than half as big. We also estimate that a plant pays almost 75% of a new worker's annual wage in the form of hiring costs, which is identified from the high observed wedge between the marginal product and wage and the high volatility of productivity that make any investments in the workforce risky. In Appendix 2.B.6, we also report time-varying estimates of adjustment costs. If anything, we find that convex adjustment costs tend to increase over time beyond what is implied by increases in the wage, pointing away from a reduction in frictions driving Indonesian growth.

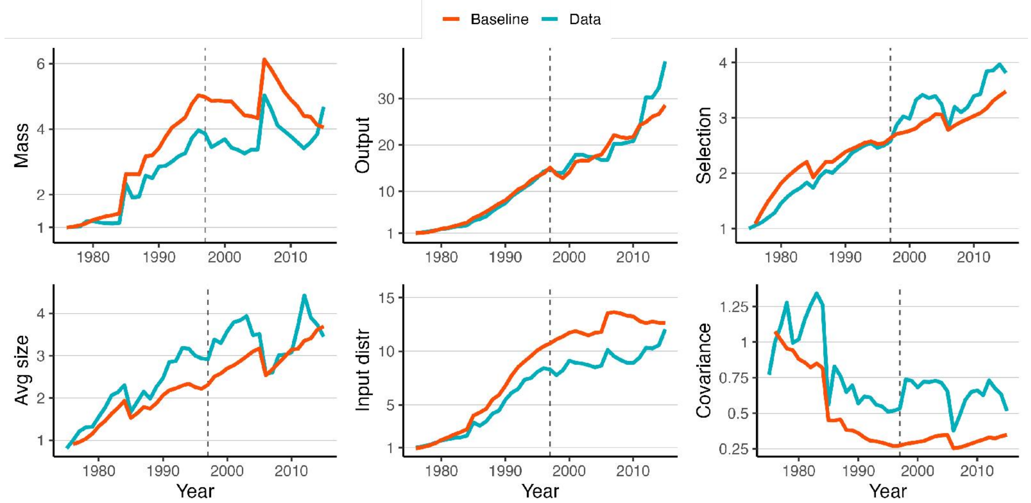
Solving the baseline model The Euler equation only identifies the ratio between the level and scale of the fixed cost distribution that determine plant exit. To separately identify the level, we solve the model and match one moment in the data and model: the mass of plants in 2015, assuming that the census is complete for that year. The estimated level and scale of the fixed cost distribution rationalize average exit rates and the low but positive correlation with underlying productivity and size.

Solving the model introduces two issues that are common to equilibrium estimation approaches. First, by requiring to solve for plants' value functions, we need to make an explicit assumption on long-run expectations of plants after the year 2015 when our data ends. We assume that after 2015, plants expect to be on a balanced growth path with manufacturing technology growing at the average rate at which it grew in the preceding ten years.[21] Given that technology in manufacturing grew strongly since 2000, this assumption implies optimistic expectations, in line with low exit and strong observed plant growth in the years prior to 2015.

The second issue for the equilibrium estimation is that enforcing revealed equilibrium prices along the estimation of the baseline economy does not guarantee that these prices actually clear markets in our model over time. We ensure consistency – informed by our specific context and data availability – by treating the observed data as correctly revealing prices, but not necessarily correctly revealing aggregate labor demand and supply. As discussed in Section 2.2.1, our data does not correctly reveal aggregate labor demand and supply due to mis- and non-reporting and as explicitly taken into account in our measure of plant exit. We thus use the

[21] We provide technical details on the BGP and how to solve for the stationarized value function in 2015 in Appendix 2.B.3. An alternative would be to solve for a continued transition towards a long-run BGP by making explicit assumptions on how all fundamentals evolve after 2015. We do not follow this approach, because it adds substantial additional computational costs while requiring similarly strong assumptions.

Figure 2.12: Baseline model fit



Notes: All graphs report results for manufacturing only and for each panel, both series are normalized to the initial level in the data. Average size is reported in efficiency units of labor. Selection (measuring average productivity), input distribution (the sum of the Cobb-Douglas aggregator of capital and labor) and covariance (the covariance between plant productivity and the share in inputs as given by the Cobb-Douglas aggregator) refer to the three corresponding terms in the growth accounting identity introduced in Section 2.

model-implied aggregate labor demand and supply for the baseline growth path and enforce the implied fundamentals that ensure market clearing for all future model-based counterfactuals. This is how our approach also ensures that counterfactuals are consistent with the baseline model economy.

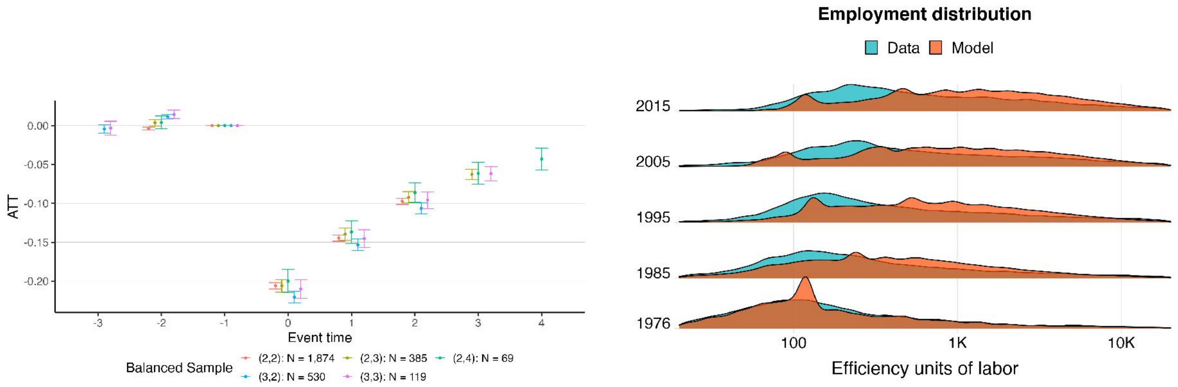
2.3.3 Evaluating model fit

To assess how well the model fits the data over time, we revisit three main results from Section 2.2. Having estimated the model on micro moments, we start out by moving from "micro to macro", evaluating the model's aggregate predictions. We then validate how well the model matches plants' dynamic input choices and changes in the entire plant distribution over time.

Figure 2.12 shows how the baseline model fits the mass of plants, aggregate output in manufacturing and the three main endogenous components that we used in Section 2.2.2 to formally decompose manufacturing growth. For completeness, we also plot the evolution of the average plant size as measured in efficiency units of labor. The model closely tracks the more than 30-fold increase in manufacturing output over time, including rapid growth in the absence of technology improvements until the Asian Financial Crisis, the decline during the crisis and the fast post-crisis growth. Overall, the model tends to slightly underestimate output at the plant level given the higher model-implied mass of plants.

The accounting identity helps to understand why the model fits the aggregate data well and where the model underperforms. First, as the main component of aggregate output growth, the model closely tracks the distribution of labor and capital across the endogenous plant distribution over time. Neither total labor demand, the total number of plants over time (apart from the first and last year) nor the distribution of inputs is hit by construction. The model captures the right degree of slow labor accumulation across the entire plant size distribution over time, which can be even more clearly seen from looking at the average plant size. Secondly,

Figure 2.13: Model validation: Event study and changes in distribution



Notes: Left graph replicates event study exactly as in Section 2 using simulated data from model growth path. Right graph gives changes in the employment distribution (using efficiency units of labor) for model versus data. For both data and model, wage bills are used and divided by (same) estimated wage path.

the model tracks well the evolution of average productivity across plants, capturing well the endogenous selection of plants over time. If anything, the model overpredicts productivity growth in the early years and underpredicts towards the end. This can be in part explained by a too fast productivity convergence implied by the estimated productivity process, stemming from frequent temporary productivity shocks that lead to overestimating productivity transitions. At last, the model also performs reasonably well on the most difficult part: the endogenous evolution of the joint distribution of productivity and inputs, as captured by the covariance term. Here, the model captures the decline in the covariance over time, as many small and productive plants enter and resources only reallocate slowly due to sizable adjustment frictions.

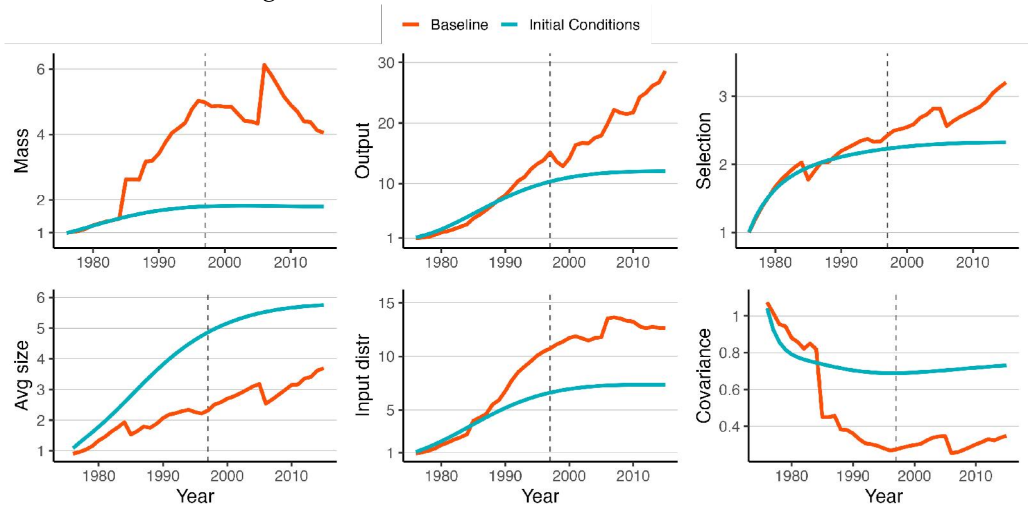
Next, we are interested in whether the model is also in line with the micro-level labor dynamics. Figure 2.13 Panel A shows the same estimated event study as in the data, using the exact same treatment definition and sample restrictions but now using simulated data from our baseline model from 1976 to 2015. The event study results align well. Both pre-trend, the exact magnitude of the treatment effect at impact and the slow recovery of the labor share follow the data. We also note that the balanced sample restrictions imposed in the data lead to very similar sample sizes in the simulated data indicating that we identify similarly selected sets of plants (which received a rare permanent productivity shock).

At last, Figure 2.13 Panel B compares model-implied versus observed changes in the entire employment distribution of plants over time. In 1976, the first year that our model predicts plant decisions, the distributions are still largely indistinguishable. Over time, the employment distribution moves strongly to the right with average employment increasing almost 4-fold and the mass of the distribution shifting from a strong left tail towards the right. The model tracks this overall change well, but slightly overpredicts the right tail. Importantly, we do not see a marked deterioration of the distribution even after 40 years of endogenous evolution.

2.4 Quantifying the drivers of aggregate growth

Using the estimated model, we now quantify the drivers of growth. Specifically, we quantify (1) the importance of initial transition growth, (2) the continuing importance of transition

Figure 2.14: Growth from initial conditions



Notes: Results for counterfactual where economy evolves only based on initial conditions (all fundamentals fixed to initial level). All graphs report results for manufacturing only. Average size is reported in efficiency units of labor. Selection (measuring average productivity), input distribution (the sum of the Cobb-Douglas aggregator of capital and labor) and covariance (the covariance between plant productivity and the share in inputs as given by the Cobb-Douglas aggregator) refer to the three corresponding terms in the growth accounting identity introduced in Section 2.

growth over the course of development, and (3) the role of policy. We present each in turn.

2.4.1 Initial conditions and the role of transition growth

We start by quantifying the importance of transition growth based on the initial economy at the onset of the Indonesian growth miracle in 1976. How much would the 1976 economy have grown purely from transition growth in the absence of any further changes in model fundamentals? For this, we start from the initial economy with the initial distribution M_{1976}, fix initial model fundamentals $\bar{\Theta}^F_{1976}$ to their value in 1976 and solve for the perfect foresight transition path (see Section 2.3), which includes solving for a counterfactual path of equilibrium wages that clears the labor market over time as the initial distribution of plants transitions towards the steady state distribution defined by $\bar{\Theta}^F_{1976}$.

Figure 2.14 highlights the resulting counterfactual growth in manufacturing. Overall output in manufacturing increases roughly 12-fold over time, accounting for 42% of the overall output gains compared to the baseline (model) economy. The reason is that young and small plants – which dominate the initial distribution and new entrants – gradually hire more workers and increase their productivity through a combination of productivity convergence and the exit of less productive plants. At the same time, entry consistently exceeds exit and the mass of plants gradually doubles over time. The increase in workers across manufacturing plants is mostly driven by the reallocation of labor from the rest of the economy: in the absence of observed changes in the aggregate labor supply and technology in the rest of the economy, the model predicts that Indonesia would have seen a manufacturing miracle with manufacturing labor and output shares reaching close to 30% over 40 years (in contrast to observed shares of less than 10%). Average plant size increases far more rapidly because in the absence of productivity improvements in the rest of the economy, aggregate wages stay almost 40% lower than in the baseline economy. Hence, initial conditions explain more than all of the increase in the average plant size over time, with cheap labor being the main driving force. Looking at changes in the entire distribution, Figure C.2 highlights that the increase in the average plant size is indeed driven by a movement in the right tail, which forms slowly because it takes time to grow large plants and the initial distribution lacks large plants, in line with the empirical evidence in Section 2.2.2.

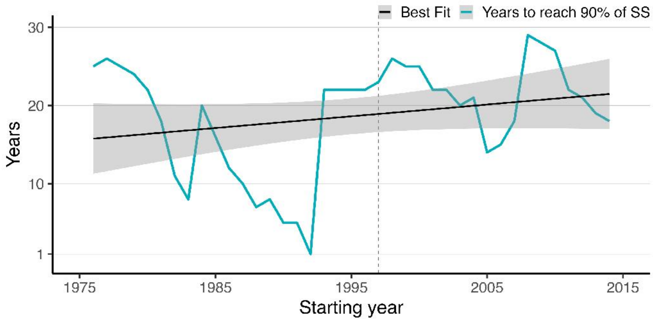
What about economy-wide effects? Aggregate output per worker increases by roughly 25% by 2015 in this counterfactual economy, explaining 5.2% of the close to 6-fold increase in aggregate output per worker observed between 1976 and 2015. However, this comparison may be unfair given that a large part of the 6-fold increase in aggregate output per worker is driven by changes in the rest of the economy, not by manufacturing. In the end, manufacturing in our data captures less than 10% of aggregate output in the economy. We thus also consider a counterfactual in which only the rest of the economy fundamentals change as observed, but manufacturing fundamentals and the initial distribution of manufacturing plants stay fixed to their values in 1976. Using this counterfactual to "purge" the effects of changes in the rest of the economy and isolate the effects of changes in manufacturing only, we find that initial transition growth accounts for all (117%) of the aggregate output per worker gains that are due to changes in manufacturing by 2015.

2.4.2 The never-ending race: Transition growth remains important

As evidenced in the previous section, even in the absence of further changes in model fundamentals, transition dynamics – due to slow labor hiring and firing and slow plant entry and exit dynamics – take decades to play out. A key question is whether the Indonesian economy runs out of transition growth over time as new changes in model fundamentals provide new potential for transition growth. We find that the quantitative answer to this question is no. We show this result by revisiting the previous exercise but instead of computing the transition path based only on the initial distribution in 1976 and initial model fundamentals $\bar{\Theta}^F_{1976}$, we compute the perfect foresight transition paths for each year between 1976 and 2015 starting from that year's initial distribution and model fundamentals $\bar{\Theta}^F_t$. This gives a total of 40 different counterfactual transition paths with their corresponding counterfactual equilibrium wage paths and plant distributions. As a measure of the transition potential, we then calculate for each transition path the number of years it takes to reach 90% of the (time-varying) steady state manufacturing output. Figure 2.15 shows that it takes the 1976 economy 26 years to come close to the steady state if fundamentals were to remain constant at $\bar{\Theta}^F_{1976}$. On average, it takes about 20 years and, importantly, the number of years to come close to the steady state does not systematically decline and – if anything – increases over time.

This race between catching up to the steady state and changes in the steady state itself can only be studied in a model that features both transition growth and changes in fundamentals and we find strong quantitative evidence that due to the combination of large and frequent changes in fundamentals and slow transition dynamics, the Indonesian economy does not get closer to its time-varying steady state. Large demographic changes and policy changes pre-1975 also provide a simple explanation for why the initial Indonesian economy in 1976 was far away from its steady state. At last, potential for transition growth may not always provide a positive force for economic growth; in fact, after 40 years of demographic changes, the mass of plants in 2015 is above its steady state and transition growth is now negative as the mass of plants

Figure 2.15: Distance to steady state over time



Notes: Years to reach 90% of the steady state manufacturing output for each transition path over 1976-2015. Each year's perfect foresight equilibrium transition path starts from that year's initial distribution and fixes fundamentals of that year over the transition. Best linear fit includes 95% CIs. Jumps are partly driven by census years in which potential entrant distributions change more strongly.

slowly declines along the transition – an important consequence of an aging population.

2.4.3 The role of policy

While the Indonesian growth experience is driven by a never-ending race of transition growth and changes in model fundamentals that induce new transition growth, the question is how government policy enters. The simple answer is: policy drives part of the changes in model fundamentals. Thus, to evaluate the effect of policy, we need to link changes in policies to changes in model fundamentals. In this section, we show how to do this by focussing on two specific but very important Indonesian government policy changes since 1975: education reform that maps to changes in human capital (and thus aggregate effective labor supply), and changes in Indonesia's foreign direct investment (FDI) policy that map to changes in the distribution of potential foreign entrants over time. In both cases, we first quantify the overall effect of changes in the specific model fundamental and then quantify the (relative) effect that can be attributed to specific policy changes. We show that overall changes in human capital were large and a necessary condition for Indonesia's manufacturing take-off, but observed education policies only explain 5% of this effect. In contrast, we find that the overall growth effects of FDI in Indonesian manufacturing were modest, but that observed changes in FDI policy explain up to 85% of its effects.

The role of cheap labor & the INPRES school construction program

What are the economy-wide and manufacturing growth effects of dramatic increases in human capital in the Indonesian economy? Over the period 1976 to 2015, our estimates suggest that human capital per worker H_t/L_t increased by 220%. To quantify the overall effects of human capital increases, we consider a counterfactual in which the Indonesian economy had not seen any human capital per worker increases over time. That is, we consider a counterfactual growth path where we start from the initial distribution in 1976 and a counterfactual path of model fundamentals with a modified path for the aggregate labor supply: $\Theta^F_{1976} \setminus \{H_t/L_t = H_{1976}/L_{1976}\}_t^\infty$. To quantify the extent to which policy contributed to the overall increases in human capital, we evaluate the effects of a particular educational policy change. Namely, we evaluate the largest school construction program in Indonesia's history and one of the largest in the world: the 1970s INPRES school construction program. The pro-

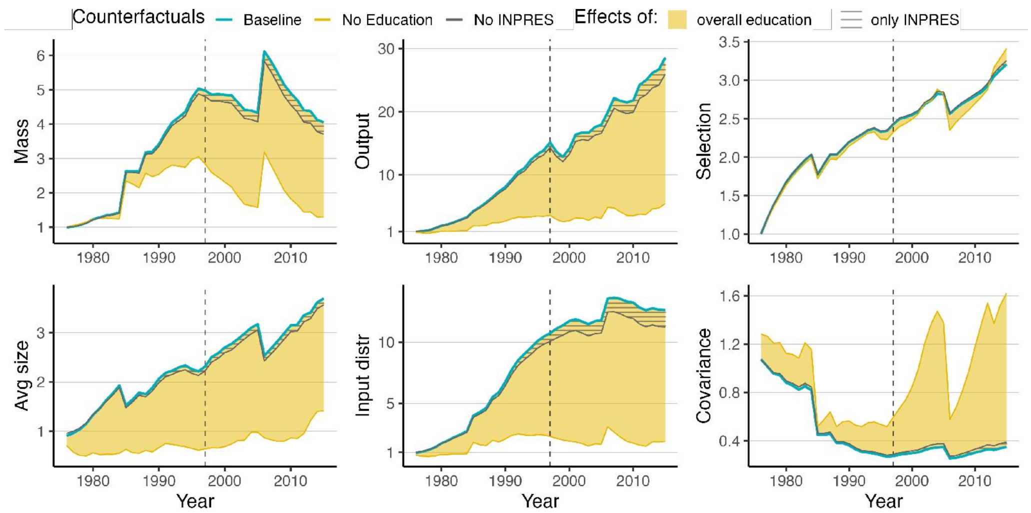


Figure 2.16: The role of education in Indonesia's manufacturing growth.

gram successfully led to increases in schooling and wages (see: Duflo 2001; Akresh, Halim, and Kleemans 2023). We assume the program only affected the Indonesian economy through its effects on human capital. To quantify the effects of the INPRES program, we consider a counterfactual in which all gains in human capital materialized except the ones that were due to the INPRES program. For this we construct a counterfactual path of aggregate human capital in the absence of the INPRES program, drawing on existing estimates on the wage effects of the program (which map to marginal changes in human capital), the known scale of the program and the increasing share of treated cohorts over time (details in Appendix 2.C.2).

Overall, we find that the estimated 220% increase in human capital per worker increased aggregate output per worker by 26.5% in 2015. The seemingly small aggregate effect is explained by output in the rest of the economy being less dependent on labor, as captured by a low estimated labor elasticity. Figure 2.16 visualizes the quantitative effects that increases in human capital had on manufacturing growth over the period 1976-2015. Manufacturing growth in Indonesia relied heavily on the cheap labor that increases in human capital brought; in the absence of this effective increase in the supply of labor ("No Education"), wages would have roughly doubled and Indonesia would not have developed a successful manufacturing sector, with manufacturing output and its employment share not even reaching 1/4 of their historical level. Average plant size would have only increased marginally over time, far less plants would have entered and far more would have exited. The INPRES program only accounts for a small share of these effects, roughly driving 5% of the aggregate output per worker gains from increases in human capital. Figure 2.16 shows that the effect on manufacturing output, the mass of producing plants and overall hiring, is approximately twice as big as the aggregate effects. This is because manufacturing is more sensitive to labor costs than the rest of the economy. Furthermore, the positive effects of the INPRES program slowly increase over time as more cohorts of Indonesians that benefited from the new-built schools enter the labor market.

The role of foreign ownership and FDI policy

Next, we look at the role of foreign direct investment in manufacturing, which we define as the foreign ownership of manufacturing plants. Foreign-owned manufacturing plants are quantitatively important, accounting for roughly 30% of manufacturing output in 2015 (see Figure C.3 in Appendix 2.C.3) and the aggregate importance of foreign ownership increased steadily since the late 1980s. FDI policy primarily affects the entry of foreign-owned plants, since ownership shares are highly persistent and most variation in foreign ownership is across, not within plants. We thus assume that FDI only affects the Indonesian growth experience through changing the distribution of potential entrants – a model fundamental that is robust to time variation in the incentives to enter – and consider counterfactual growth paths in which we only change the path of potential entrant distributions. Again, we want to separately quantify the effects of FDI and the relative effect that changes in observed FDI policy had on FDI. For this, we separate the distribution of potential entrants at any point in time into the distribution of potential foreign entrants and potential domestic entrants enforcing model-consistent entry decisions. We then construct a counterfactual path of potential entrant distributions without foreign entrants. To capture the effect of policy, we consider important regulatory changes in FDI policy in 1987. Specifically, we exploit variation in potential foreign entrant distributions right before and after the reform to measure the effect of policy and use the estimated effect to construct a counterfactual path of potential foreign entrant distributions in the absence of the FDI policy change (details in Appendix 2.C.3).

We find that FDI helped manufacturing growth, but did not play a transformative role. Specifically, the entry of foreign-owned manufacturing plants explains 7.5% of the aggregate output per worker gains due to manufacturing growth and we estimate that manufacturing output and the manufacturing employment share would be 8% lower in 2015 in the absence of FDI. The reason for this rather small effect is that given a high estimated supply of domestic potential entrants, the downward pressure on labor demand and wages due to the disappearance of foreign entrants leads to an elastic response of domestic entry in general equilibrium that mitigates some of the negative effects of losing FDI. In contrast to the case of education policy, we find that *changes* in FDI policy potentially explain most of the overall growth effects from FDI. Specifically, changes in FDI policy potentially explain a four- to five-fold increase in potential entry and these changes in FDI policy in turn explain 85% of the overall growth effects from FDI.

2.5 Conclusion

This paper studied the drivers of growth miracles. Building on 40 years of plant-level manufacturing panel data for Indonesia, we motivated a model in which rapid growth is driven by a combination of transition growth and changes in fundamentals that are dominated by worker and plant demographics. We showed how to tractably estimate this model on the observed growth path using standard plant-level data and without assuming that the observed economy is at a steady state at any point in time. We found that transition growth is key: 42% of the observed manufacturing output growth is simply explained by initial conditions in 1975 – dominated by young and small manufacturing plants – providing ample opportunities for catch-up growth. Transition growth also does not become less important over time because important demographic changes in the economy induce further potential for transition growth.

Since our model and estimation framework maps directly to observed time-varying aggregate growth and its micro-level drivers of endogenous changes in the distribution of plants, it is particularly suited to study the dynamic growth effects of observed policy. This link is important not only to better validate macroeconomic models of growth, but also to study the aggregate growth and general equilibrium effects of policy – effects that are rarely identified in micro-empirical policy evaluations. In this paper, we only started looking at this by showing how to use the estimated model to evaluate the dynamic growth effects of two important Indonesian policies: education reform and changes in FDI policy. Based on our results, a somber conclusion – partly resonating related work on the Indian growth miracle (Bollard, Klenow, and Sharma 2013) – is that observed policy mattered less for growth than we might think. Instead, we find that Indonesian growth was mostly driven by structural forces related to demographics. This does not mean that policy necessarily plays no role. In fact, in Appendix 2.C.4 we consider two sets of reduced-form policies – a reduction in (convex) labor adjustment frictions and an increase in annual technology growth – that both would have doubled Indonesian manufacturing output by 2015, an even more remarkable manufacturing miracle closer to experiences in countries such as China and Malaysia. Future research should further unpack what drives changes in adjustment costs, technology growth or the pool of potential entering plants and link these closer to policy.

2.6 References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica* 83 (6): 2411–51.

Akcigit, Ufuk, Harun Alp, and Michael Peters. 2021. "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries." *American Economic Review* 111 (1): 231–75.

Akresh, Richard, Daniel Halim, and Marieke Kleemans. 2023. "Long-Term and Intergenerational Effects of Education: Evidence from School Construction in Indonesia." *The Economic Journal* 133 (650): 582–612.

Alessandria, George, Costas Arkolakis, and Kim J. Ruhl. 2021. "Firm Dynamics and Trade." *Annual Review of Economics* 13 (1): 253–80.

Allen, Treb, and Dave Donaldson. 2020. "Persistence and Path Dependence in the Spatial Economy." National Bureau of Economic Research.

Asturias, Jose, Sewon Hur, Timothy J. Kehoe, and Kim J. Ruhl. 2023. "Firm Entry and Exit and Aggregate Growth." *American Economic Journal: Macroeconomics*.

Atkeson, Andrew, and Ariel Tom'as Burstein. 2010. "Innovation, Firm Dynamics, and International Trade." *Journal of Political Economy* 118 (3): 433–84. https://doi.org/10.1086/653690.

Atkin, David, Amit K. Khandelwal, and Adam Osman. 2019. "Measuring Productivity: Lessons from Tailored Surveys and Productivity Benchmarking." In *AEA Papers and Proceedings*, 109:444–49. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203. https://www.aeaweb.org/articles?id=10.1257/pandp.20191005.

Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin. 2007. "Estimating Dynamic Models of Imperfect Competition." *Econometrica* 75 (5): 1331–70.

Basri, M. Chatib, Mayara Felix, Rema Hanna, and Benjamin A. Olken. 2021. "Tax Administration Versus Tax Rates: Evidence from Corporate Taxation in Indonesia." *American Economic Review* 111 (12): 3827–71. https://doi.org/10.1257/aer.20201237.

Bernstein, Shai, Emanuele Colonnelli, Davide Malacrino, and Tim McQuade. 2022. "Who Creates New Firms When Local Opportunities Arise?" *Journal of Financial Economics* 143 (1): 107–30.

Bilal, Adrien, Niklas Engbom, Simon Mongey, and Giovanni L. Violante. 2022. "Firm and Worker Dynamics in a Frictional Labor Market." *Econometrica* 90 (4): 1425–62.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *The Quarterly Journal of Economics* 128 (1): 1–51.

Bloom, Nicholas, Aprajit Mahajan, David McKenzie, and John Roberts. 2020. "Do Management Interventions Last? Evidence from India." *American Economic Journal: Applied Economics* 12 (2): 198–219.

Bollard, Albert, Peter J. Klenow, and Gunjan Sharma. 2013. "India's Mysterious Manufacturing Miracle." *Review of Economic Dynamics* 16 (1): 59–85.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch. 2021. "Some Unpleasant Markup Arithmetic: Production Function Elasticities and Their Estimation from Production Data." *Journal of Monetary Economics* 121: 1–14.

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang. 2012. "Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing." *Journal of Development Economics* 97 (2): 339–51.

Buera, Francisco, Joseph Kaboski, and Robert Townsend. 2023. "From Micro to Macro Development." *Journal of Economic Literature* 61 (2): 471–503.

Buera, Francisco, and Yongseok Shin. 2013. "Financial Frictions and the Persistence of History: A Quantitative Exploration." *Journal of Political Economy* 121 (2): 221–72.

Cali, Massimiliano, Marco Le Moglie, and Giorgio Presidente. 2021. "Gain Without Pain? Non-Tariff Measures, Plants' Productivity and Markups."

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro. 2019. "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock." *Econometrica* 87 (3): 741–835.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–230.

Chen, Zhang. 2022. "Economic Growth and the Rise of Large Firms." *Job Market Paper*. https://zhangc-econ.github.io/files/TailGrowth.pdf.

Choi, Jaedo, Andrei A. Levchenko, Dimitrije Ruzic, and Younghun Shim. 2023. "Superstars or Supervillains? Large Firms in the South Korean Growth Miracle." https://www.freit.org/EIIT/2023/selected/choi.pdf.

Clementi, Gian Luca, and Berardino Palazzo. 2016. "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations." *American Economic Journal: Macroeconomics* 8 (3): 1–41.

Cooper, Russell, Guan Gong, and Ping Yan. 2018. "Costly Labour Adjustment: General Equilibrium Effects of China's Employment Regulations and Financial Reforms." *The Economic Journal* 128 (613): 1879–1922.

Cooper, Russell, John Haltiwanger, and Jonathan L. Willis. 2015. "Dynamics of Labor Demand: Evidence from Plant-Level Observations and Aggregate Implications." *Research in Economics* 69 (1): 37–50.

Coşar, A. Kerem, Nezih Guner, and James Tybout. 2016. "Firm Dynamics, Job Turnover, and Wage Distributions in an Open Economy." *American Economic Review* 106 (3): 625–63.

Demirer, Mert. 2020. "Production Function Estimation with Factor-Augmenting Technology: An Application to Markups." *UPenn. Job Market Paper*.

Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 91 (4): 795–813.

———. 2004. "The Medium Run Effects of Educational Expansion: Evidence from a Large School Construction Program in Indonesia." *Journal of Development Economics* 74 (1): 163–97.

Eslava, Marcela, and John C. Haltiwanger. 2020. "The Life-Cycle Growth of Plants: The Role of Productivity, Demand and Wedges." National Bureau of Economic Research.

Eslava, Marcela, John Haltiwanger, and Alvaro Pinzon. 2022. "Job Creation in Colombia Versus the USA: Up or Out Dynamics Meet the Life Cycle of Plants." *Economica* 89 (355): 511–39.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2016. "The Slow Growth of New Plants: Learning about Demand?" *Economica* 83 (329): 91–129.

Gandhi, Amit, Salvador Navarro, and David Rivers. 2017. "How Heterogeneous Is Productivity? A Comparison of Gross Output and Value Added." *Journal of Political Economy*.

Gathen, Jonas. 2021. "Government Jobs for Talents, Cronies or Sale? Estimating Government Worker Skills with an Application to Indonesia." *Working Paper*.

Gillis, Malcolm. 1985. "Micro and Macroeconomics of Tax Reform: Indonesia." *Journal of*

Development Economics 19 (3): 221–54.

Gopinath, Gita, Şebnem Kalemli-Özcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. 2017. "Capital Allocation and Productivity in South Europe." *The Quarterly Journal of Economics* 132 (4): 1915–67.

Hansen, Lars Peter, and Kenneth J. Singleton. 1982. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica: Journal of the Econometric Society*, 1269–86. https://www.jstor.org/stable/1911873.

Hastie, Trevor J. 2017. "Generalized Additive Models." In *Statistical Models in S*, 249–307. Routledge.

Hausmann, Ricardo, Lant Pritchett, and Dani Rodrik. 2005. "Growth Accelerations." *Journal of Economic Growth* 10: 303–29.

Hicks, Joan Hamory, Marieke Kleemans, Nicholas Y. Li, and Edward Miguel. 2017. "Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata." National Bureau of Economic Research.

Hill, Hal. 2000. *The Indonesian Economy*. Cambridge University Press.

Hopenhayn, Hugo. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, 1127–50.

———. 2014. "Firms, Misallocation, and Aggregate Productivity: A Review." *Annual Review of Economics* 6 (1): 735–70.

Hopenhayn, Hugo, and Richard Rogerson. 1993. "Job Turnover and Policy Evaluation: A General Equilibrium Analysis." *Journal of Political Economy* 101 (5): 915–38.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein. 2016. "Quantifying the Sources of Firm Heterogeneity." *The Quarterly Journal of Economics* 131 (3): 1291–1364.

Hotz, V Joseph, and Robert A Miller. 1993. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *The Review of Economic Studies* 60 (3): 497–529.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *The Quarterly Journal of Economics* 124 (4): 1403–48.

———. 2014. "The Life Cycle of Plants in India and Mexico." *The Quarterly Journal of Economics* 129 (3): 1035–84.

Hsieh, Chang-Tai, and Benjamin A. Olken. 2014. "The Missing" Missing Middle"." *Journal of Economic Perspectives* 28 (3): 89–108.

Humlum, Anders. 2022. "Robot Adoption and Labor Market Dynamics."

Karahan, Fatih, Benjamin Pugsley, and Ayşegül Şahin. 2019. "Demographic Origins of the Startup Deficit." National Bureau of Economic Research.

Kleinman, Benny, Ernest Liu, and Stephen J. Redding. 2023. "Dynamic Spatial General Equilibrium." *Econometrica* 91 (2): 385–424.

Klette, Tor Jakob, and Samuel Kortum. 2004. "Innovating Firms and Aggregate Innovation." *Journal of Political Economy* 112 (5): 986–1018. https://doi.org/10.1086/422563.

Kruse, Hagen, Emmanuel Mensah, Kunal Sen, and Gaaitzen de Vries. 2023. "A Manufacturing (Re) Naissance? Industrialization in the Developing World." *IMF Economic Review* 71 (2): 439–73.

Lanteri, Andrea, Pamela Medina, and Eugene Tan. 2023. "Capital-Reallocation Frictions and Trade Shocks." *American Economic Journal: Macroeconomics* 15 (2): 190–228.

Lent, G., and P. Ojha. 1969. "Sales Taxes in Countries of the Far East." *IMF Staff Papers* 16.

Liang, James, Hui Wang, and Edward P. Lazear. 2018. "Demographics and Entrepreneurship." *Journal of Political Economy* 126 (S1): S140–96.

Luttmer, Erzo GJ. 2007. “Selection, Growth, and the Size Distribution of Firms.” *The Quarterly Journal of Economics* 122 (3): 1103–44.

McCaig, Brian, and Nina Pavcnik. 2021. “Entry and Exit of Informal Firms and Development.” *IMF Economic Review* 69: 540–75.

McKenzie, David, and Anna Luisa Paffhausen. 2019. “Small Firm Death in Developing Countries.” *Review of Economics and Statistics* 101 (4): 645–57.

Midrigan, Virgiliu, and Daniel Yi Xu. 2014. “Finance and Misallocation: Evidence from Plant-Level Data.” *American Economic Review* 104 (2): 422–58.

Moll, Benjamin. 2014. “Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?” *American Economic Review* 104 (10): 3186–3221.

Olley, G. Steven, and Ariel Pakes. 1996. “The Dynamics of Productivity in the Telecommunications Equipment Industry.” *Econometrica*, 1263–97.

Ottonello, Pablo, and Thomas Winberry. 2023. “Investment, Innovation, and Financial Frictions.”

Perla, Jesse, Christopher Tonetti, and Michael E. Waugh. 2021. “Equilibrium Technology Diffusion, Trade, and Growth.” *American Economic Review* 111 (1): 73–128.

Peters, Michael. 2020. “Heterogeneous Markups, Growth, and Endogenous Misallocation.” *Econometrica* 88 (5): 2037–73. https://doi.org/10.3982/ECTA15565.

Poschke, Markus. 2013. “‘Entrepreneurs Out of Necessity’: A Snapshot.” *Applied Economics Letters* 20 (7): 658–63.

———. 2018. “The Firm Size Distribution Across Countries and Skill-Biased Change in Entrepreneurial Technology.” *American Economic Journal: Macroeconomics* 10 (3): 1–41.

Restuccia, Diego, and Pedro Bento. 2015. “Misallocation, Establishment Size, and Productivity.” In *2015 Meeting Papers*. 1391. Society for Economic Dynamics.

Restuccia, Diego, and Richard Rogerson. 2017. “The Causes and Costs of Misallocation.” *Journal of Economic Perspectives* 31 (3): 151–74.

Romer, Paul M. 1990. “Endogenous Technological Change.” *Journal of Political Economy* 98 (5, Part 2): S71–102.

Ruggieri, Alessandro. 2022. “Trade and Labour Market Institutions: A Tale of Two Liberalizations.” University of Nottingham, GEP.

Ruiz-García, J. C. 2019. “Financial Frictions, Firm Dynamics and the Aggregate Economy: Insights from Richer Productivity Processes.” Working Paper.

Schoar, Antoinette. 2010. “The Divide Between Subsistence and Transformational Entrepreneurship.” *Innovation Policy and the Economy* 10 (1): 57–81.

Timmer, Marcel, Gaaitzen J. de Vries, and Klaas De Vries. 2015. “Patterns of Structural Change in Developing Countries.” In *Routledge Handbook of Industry and Development*, 79–97. Routledge.

Ulyssea, Gabriel. 2020. “Informality: Causes and Consequences for Development.” *Annual Review of Economics* 12 (1): 525–46.

Van Biesebroeck, Johannes. 2005. “Firm Size Matters: Growth and Productivity Growth in African Manufacturing.” *Economic Development and Cultural Change* 53 (3): 545–83.

World Bank. 2023. *Poverty and Inequality Platform (Version 20230328_2017) [Data Set]. World Bank Group. Www.pip.worldbank.org. Accessed the 15th September, 2023.*

Appendices

2.A Data and Empirical Evidence

2.A.1 Data cleaning details

In the following subsection, we describe in detail the data cleaning steps we take to ensure that the data is consistent over time and that results are not driven by different forms of measurement error. The main data cleaning steps relate to cleaning plant-level labor (wage bill and workers) and output (value added) over time. We further discuss how we deal with "dynamic" outliers (e.g. unrealistic within-plant jumps in value added or the wage bill) and observations with extreme labor shares that pose numerous problems for the estimation and computation of the model. Besides these mentioned cleaning steps, we also drop a few clear outliers, such as when the plant ID is misreported or missing or when magnitudes of multiple reported variables are impossible. (Give details on final cleaning: How does raw data differ from cleaned data). At last, we report details on how we clean the capital series, industry codes and measure plant entry and exit.

Cleaning labor and the labor wage bill

The manufacturing census consistently reports a plant's total number of workers (including paid and unpaid) as well as separately the number of paid versus unpaid workers and the number of production and non-production (including managerial) workers. The main cleaning step we apply to ensure consistency over time is to drop all plants with less than 20 total workers (which is enforced by BPS starting in 1990, but not before), drop plants that report zero paid workers or that report more paid workers than total workers. This step drops slightly less than 2% of plant-year observations with dropped observations concentrated before 1990. We also identify a bunching at 99 workers in the years 2013-2015 (roughly 3-4% of plants), which we interpret as true bunching driven by actual policy changes and thus do not correct.[22]

For the structural model, we build on a plant's reported total wage bill. This variable is the sum of the total wage bill for other workers and for production workers. In principle, it includes all payments to labor, including in-kind transfers, overtime pay, bonuses and social contributions (e.g. pension and accident allowances). Since the survey asks about current workers and doesn't separately ask about severance pay, we treat the reported wage bill as excluding severance pay.

[22] For example, Indonesia introduced an occupational safety and health regulation in 2012 that was targeted at manufacturing firms and mandates any workplace with more than 100 workers to implement additional work safety measures.

We take three main cleaning steps for the reported wage bill. First, we correct systematic misreporting for the year 2011. Looking at the evolution of the distribution of the total wage bill across years, we find that 2011 is the only clear outlier with the only bimodal wage distribution across all years. In 2011, the bottom 20% of observations show exactly the same per worker wage at an unrealistically low level that is below the bottom 1% of observations in 2010 and 2012 and roughly 50 times lower than the average minimum wage in 2011. Given that these are well-defined misentries, the remainder of the distribution is well-behaved and we observe most plants before and after 2011, we opt for imputing misentries using within-plant averages across 2010 to 2012, enforcing linear growth for 2011.

Second, we correct for misreporting in the wages for non-production workers. Non-production workers account for roughly 16.5% of overall employment in our data. However, 17% of plant-year observations reportedly employ zero non-production workers. This also means that plants may not always report all managerial workers, which is likely if the managerial staff partly owns the establishment (and is thus not formally employed). To the extent that all payments to managerial workers should be counted as labor costs, the Indonesian data may significantly underestimate labor costs. We cannot correct for this form of underreporting. However, we can correct for the following: About 10% more plant-year observations report zero wage payments to non-production workers than plant-year observations reporting zero non-production workers. That is, there are plants that report employing non-production workers, but paying them no wages. This could be in part due to some plants reporting managerial staff who receive remuneration other than wages or due to plants simply not reporting non-production worker wages. In any case, we think it is better to impute wages here, using plants' reported wages for production workers and the average year-specific pay gap across production and non-production workers for plants that report both wages. We find that wage premia for non-production workers were around 90% in 1975 and declined to around 20% in 2015. In the end, the overall importance of this correction is small, because the correction only applies to a small number of observations.

In the third and last cleaning step, we correct the total reported wage bill across periods where the exact questions and components of the total wage bill changed. While plants were asked consistently to report total payments including cash and in-kind wages, pensions and other social contributions between 1975-1995, survey questions changed most notably between 2001-2003, 2004-2010 and 2011-2015. Looking at changes in the distribution of reported total wage bills, the 2004-2010 period appears to be the most problematic period in which total wages are systematically underreported vis-a-vis the other periods.

We correct for changes in the measurement of the total wage bill over the period 2004-2010 by exploiting within-plant changes in reported wage bills across changes in the measurement period (from 2003 to 2004 and from 2010 to 2011) and further utilizing information across different types of workers and the reporting of the number of workers by type. Plants i report x^j_{it}, the total reported wage bill for worker type j in period t. Specifically, we assume that: $x^j_{itm} = w_t h^j_{it} \epsilon^j_{it} \tau^j_m$, where w_t is the wage (in line with our model), h^j_{it} are plant-worker-type-specific efficiency units of labor, ϵ^j_{it} captures idiosyncratic measurement error and τ^j_m is systematic underreporting that is constant within a worker-type and within the measurement period m where the same questions to elicit the worker-type-specific wage bill are asked. We assume that $\tau^j_m \in (0, 1]$ for the period 2004-2010 and unity otherwise.

Our approach separately identifies wages from the measurement error τ_m^j across time and thus allows to correct for an important part of measurement error that leads to the underestimation of the total wage bill and a time inconsistent measure of the wage bill. For separate identification, we assume that over two consecutive periods, plants use the same average human capital within types of worker: $\frac{h_{it}^j}{l_{it}^j} = \frac{h_{it+1}^j}{l_{it+1}^j} = \alpha_i^j$. This allows some plants to specialize on high productive production workers or other plants to specialize on low productive managerial staff, but restricts changes in the average human capital within plant-worker-types. As long as the plant-specific number of workers by type are reported either without measurement error or with constant plant-worker-type measurement error, the assumption allows to identify:

$$\mathbb{E}_i \frac{\frac{x_{i,t+1,m}^j}{l_{i,t+1,m}^j}}{\frac{x_{i,t,m}^j}{l_{i,t,m}^j}} \equiv \mathbb{E}_i \frac{\tilde{x}_{i,t+1,m}^j}{\tilde{x}_{i,t,m}^j} = \frac{w_{t+1}}{w_t} \mathbb{E}_i \frac{\epsilon_{it+1}^j}{\epsilon_{it}^j}$$

Within measurement periods m, one can show that under realistic magnitudes for the measurement error, the following holds:[23]

$$\mathbb{E}_i \frac{\tilde{x}_{i,t+1,m}^j}{\tilde{x}_{i,t,m}^j} \approx \frac{w_{t+1}}{w_t}$$

Across measurement periods, separate identification of the change in measurement and the wage is impossible without further assumptions. To see this, write:

$$\mathbb{E}_i \frac{\tilde{x}_{i,t+1,m'}^j}{\tilde{x}_{i,t,m}^j} = \frac{w_{t+1}}{w_t} \mathbb{E}_i \frac{\epsilon_{it+1}^j \tau_{m'}^j}{\epsilon_{it}^j \tau_m^j} \approx \frac{w_{t+1}}{w_t} \frac{\mathbb{E}_i \epsilon_{it+1}^j \tau_{m'}^j}{\mathbb{E}_i \epsilon_{it}^j \tau_m^j} = \frac{w_{t+1}}{w_t} \frac{\mathbb{E}_i \tau_{m'}^j}{\mathbb{E}_i \tau_m^j}$$

where we have made use of the same approximation as above. Even then, changes in the measurement error across measurement periods cannot be separately identified from wage changes. To solve this issue, we interpolate wages from wage growth in the previous period and the next period (for which measurement does not change), assuming that wages grow smoothly over time.

In our case, we set $\tau_m^j = 1$ for all measurement periods except the period 2004-2011. To identify $\tau_{m'}^j$ for 2004-2011, we are now actually over-identified, because we can identify the measurement error from variation between 2003-2004 or from 2010-2011. We choose to use 2003-2004 because 2010-2011 featured a change in the minimum wage in Indonesia, which partly explains a large increase in plants' total wage bills and we do not know how to separate this change from a change in the measurement. Following this approach, we find that $\tau_{m'}^j \approx 0.94$, similar when restricting to production workers only or when looking at all workers. We en-

[23] Specifically, a first-order Taylor series approximation around the mean of the measurement errors gives: $\frac{w_{t+1}}{w_t} \mathbb{E}_i \frac{\epsilon_{it+1}^j}{\epsilon_{it}^j} \approx \frac{w_{t+1}}{w_t} \frac{\mathbb{E}_i \epsilon_{it+1}^j}{\mathbb{E}_i \epsilon_{it}^j}$. With a second-order Taylor series approximation, we get: $\mathbb{E}_i \frac{\epsilon_{it+1}^j}{\epsilon_{it}^j} \approx \frac{\mathbb{E}_i \epsilon_{it+1}^j}{\mathbb{E}_i \epsilon_{it}^j} \left[1 - \frac{Cov_i(\epsilon_{it+1}^j, \epsilon_{it}^j)}{\mathbb{E}_i \epsilon_{it+1}^j \mathbb{E}_i \epsilon_{it}^j} + \frac{Var_i(\epsilon_{it}^j)}{\left[\mathbb{E}_i \epsilon_{it}^j\right]^2} \right]$. Plugging in realistic measurement error, the 2nd-order correction is very small. For example, if measurement error within plants is positively correlated (which is the likely case), then the two correction terms go in opposite directions. Furthermore, both the covariance term and the variance term are close to zero for reasonable magnitudes of measurement error.

force the production worker correction across all plants for the measurement period 2004-2011 (which means that their wage bills get upward corrected by $1/\tau^j_{m'}$, a correction of roughly 6%).

Cleaning output / value-added

Throughout the paper, we use a consistent definition of value-added output. This definition coincides with how the Indonesian statistical agency (BPS) constructed value-added output for some, but not all years. Specifically,

$$\text{Value-added}_{it} \equiv \text{Gross income}_{it} - \text{Intermediates}_{it}$$

where $\text{Gross income}_{it} \equiv \text{Gross sales}_{it} + \text{electricity sales}_{it} + \text{revenue from industrial services}_{it} + \text{other income}_{it} + \Delta\text{value of semi-finished products}_{it}$ and $\text{Intermediates}_{it} \equiv \text{Raw materials}_{it} + \text{Total fuel/electricity expenditures}_{it} + \text{Other expenses}_{it}$. All inputs in the accounting identities are reported in their current values of Rupiah, which we deflate to 2010 constant Rupiah based on the aggregate CPI. We start by dropping observations with missing or negative gross income, which are less than 0.5% of observations.

Next, we construct a time-consistent measure of intermediates. The main issue is that intermediates are likely underreported since the survey only asks for specific categories of expenditures and intermediate expenditures have likely become more complex over time, leading plants to underreport parts of their expenditures. This leads to an overestimate of value-added output and an underestimate of capital and labor cost shares at the plant-level. We correct intermediate expenditures in two steps.

In the first step, we look at one main expenditure category of intermediate inputs for which we know that time inconsistency is an issue. Specifically, $\text{Other expenses}_{it}$ are reported inconsistently over time because not all components of other expenses are enumerated in every year. In the following, we describe the components of $\text{Other expenses}_{it}$ and how we impute them consistently across plants over time. In the years with the most detailed survey questions, $\text{Other expenses}_{it}$ (X_{it} in short) are the sum of three components (indexed by c): (1) expenses for other goods (consisting of packaging, spare parts and stationary), (2) manufacturing services, repair and maintenance, and (3) remaining other expenses (with detailed subcomponents for some years). We improve the measure of intermediate expenditures by imputing these three subcomponents in cases where they are missing. Similar to the components of the labor wage bill, we deal with underreporting of other expenses by exploiting within-plant differences in reporting around years with changes in survey questions. We separately impute missing fractions of each of the three components of other expenses, using further information on subcategories j within components c, bringing all series to the most complete level of reporting in the years 2006 and 1996/1997.

Specifically, we assume that $\forall c, j: \quad X_{icjt} = \alpha_{icj} Y_{it}$. That is, any other expense category (or subcategory) is a plant-subcategory-specific fraction of gross income Y_{it}. Since expenditures for specific subcategories are systematically missing in some years, but gross income Y_{it} is reported for all years, we impute complete missing subcategory expenditures as follows: For plants that we observe across different measurement periods, we impute their expenditure shares from average within-plant expenditure shares around the time of missing. For example, expenses for other goods are missing between 1998 and 2005, which we impute using the plant average of plant-specific expenditure shares in 1997 and 2006 together with plant-year-specific gross income Y_{it}. This ensures within-plant consistency and allows for plant-category-year-specific variation in expenditures. For plants for which we do not observe expenditures in other years, we use the aggregate category-specific expenditure share around the time of missing.

On top of this, we correct reported expenditures for the remaining other expenditures for the period 1975-1984 in which the reported series is clearly underreported in comparison to post 1985. For this correction, we again proceed separately for plants that we observe across measurement periods and for plants that we do not, using either the average within-plant difference in reported ratios or the ratio of aggregate expenditure shares across the two measurement periods as correction factors.

Overall, this first step of cleaning intermediate inputs ensures more time consistency, but does not have a sizable effect on overall intermediate expenditures. In the second step of correcting intermediate expenditures, we deal with the sizable remaining decline in the intermediate expenditure share across plants over time. For example, the aggregate intermediate expenditure share declines by more than 10 percentage points from roughly 0.65 in 1980 to 0.525 in 2015 (mostly driven by a decline in the raw material input share). We expect that a major part of this decline is in fact measurement error. One simple reason could be that plants use more processed intermediate inputs, which they do not fully report as "raw materials". To distinguish this driver from industrial composition effects (e.g. industries relying on intermediate inputs declining in relative importance over time), we construct the following correction: We regress $log(\phi_{ijt}/(1-\phi_{ijt})) = \alpha_j + \alpha_t + \epsilon_{ijt}$ where ϕ_{ijt} is the intermediate expenditure share of plant i in 5-digit industry j at time t. We use the log odds ratio to ensure that any correction we implement gives expenditure shares that are bounded between zero and one. α_j and α_t capture industry and time fixed effects. We interpret α_t as our time-varying bias term, using α_{1975} as the normalization factor (for which the bias is zero). Controlling for industry fixed effects ensures that the bias terms do not capture variation in intermediate expenditure shares from changes in the industrial composition. Corrected intermediate input shares are then given by $\tilde{\phi}_{ijt} = exp(\alpha_j + \alpha_{1975} + \epsilon_{ijt})/(1 + exp(\alpha_j + \alpha_{1975} + \epsilon_{ijt}))$. The correction maintains plant-level variation in intermediate expenditure shares and delivers both within-plant and aggregate time consistency. To ensure that the regression is well-specified, we initially drop observations with non-positive intermediate inputs or value added as well as observations with missing value added. This drops less than 2% of observations. After the correction, we recompute intermediate expenditures and value added.

Identifying problematic outliers: jumps and extreme labor shares

The last important cleaning step we take, is to identify problematic outliers that are likely misentries and would have an outsized role on the model estimation and inference.

We start with "dynamic" outliers by which we refer observations that are outliers within the time series of an individual plant. We treat a plant-year observation as a dynamic outlier if the total wage bill or value added output series (X_{it} in the following) of an individual plant sees a sizable one-time jump after which it reverts directly back. We consider two different measures: the year-to-year within-plant change $\frac{X_{it}}{X_{it-1}}$ and the year-to-year within-plant quantile difference $q_t(X_{it}) - q_{t-1}(X_{it-1})$. For both measures, we first identify a potential outlier if any of the two measures is below the 10th or above the 90th within-year percentile for the respective measure. For example, we classify the following observation a potential outlier: Between 1993 and 1994, a plant's quantile of value added output changed by more than the 90th percentile of this year's distribution of value added output quantile changes. We classify any potential outlier as an actual outlier only if the following year is also identified as a potential outlier whose change goes in the opposite direction.[24] This ensures to identify jumps, while the initial and final level do not need to coincide, allowing for plant-, time- and variable-specific drifts. Big one time changes are explicitly not counted as outliers, treating them as true shocks. We drop any dynamic outlier that we detect through this procedure. Note that this procedure also identifies observations that change back and forth multiple times in a row as outliers as long as their changes are very large. In total, this procedure drops almost 10% of plant-year observations and roughly 10% of total reported value added.

The second and last cleaning step we take is to drop observations with extreme labor shares. These are observations with reported labor shares below 5% and above 500% (roughly differing from the median labor share of 50% by a factor of 10). Extreme labor shares are likely a combination of overreported value added and underreported wage bills and these observations have a sizable impact on aggregates. They make up roughly 3% of observations, but account for 41% of total reported value added. While we think that many of these plant-year observations with large value added are correctly classified as being "granular" in their importance for output (e.g. many of these plants consistently report being large over time), their exact value added output and wage bills are likely mismeasured.

Cleaning capital

For cleaning plants' self-reported capital stock, we draw on the cleaning steps in Cali, Le Moglie, and Presidente (2021), which is the most thorough attempt at cleaning the Indonesian manufacturing plant capital series that we are aware of. The cleaning steps draw in part on the perpetual inventory method (PIM). Details can be found in Cali, Le Moglie, and Presidente (2021).

Cleaning industry codes

Industry classifications changed over time, starting with ISIC 2 in 1975 and moving to ISIC 3, ISIC 3.1 and ISIC 4 by the end of our data period. For harmonization, we start by fixing a plant's first reported 5-digit industry (the most disaggregated level reported). While plants may reasonably change industries over time, we opt for fixing industries to have a time-consistent plant-level measure of industry. We then build backward correspondences at the 2-, 3-, 4- and 5-digit industries respectively using within-plant changes in industry classifications across changes in classification systems (that is, correspondences that map from later year classifications to earlier ones). For plants that enter later, we enforce these within-plant correspondences. In the case of one-to-many mappings (e.g. the same industry code in ISIC 3.1 maps to different codes in ISIC 3), we enforce the most common one. Note that this only matters for plants that are not observed previously. In the case of no linking (e.g. a plant enters in ISIC 4 with a code that has no observed backward linkage), we check codes manually and use official crosswalks. In the cases where we think that industries are truly "new", we simply

[24] We do this separately by measure so the identification of an outlier is based on within-measure changes only.

create a new industry code. In the end, we are left with 140 unique 5-digit industry codes, 120 of which are initial ISIC 2 codes and 20 codes are new industries. At the 2-digit level, we have 9 different industries.

Plant entry and exit

In principle, classifying entry and exit should be straightforward: whenever a plant with a new panel identifier enters our panel, we would record this as plant entry and whenever that plant reports in t but does not report in $t+1$, we would record the plant as having exited at the end of t. In practice, this classification would inflate plant entry and exit because occasional non-reporting is common. This is because of actual non-reporting and – as described above – because we explicitly drop plant-year observations with misreported entries. We thus only classify a plant as having exited if we do not observe reporting by the plant at any future time period. Similarly for entry, we only count the plant as entering if it is the first time the plant identifier has entered the panel. This difference is quantitatively important: the unconditional exit rate drops almost by half from around 14% to 7.9% if we follow our classification. As we discuss in the main text, 7.9% is close to other exit rates reported for India, Mexico and the US (e.g. Hsieh & Klenow 2014).

Where and how does this matter and could this new classification bias results differentially over time? Throughout the analyses, we mostly draw on within-plant changes that are robust to compositional changes due to entry and exit. In the structural model, we explicitly continue modeling plants that are non-reporting but non-exiting, correcting for (some forms of) differential non-reporting over the state space. Bias does arise from non-reporting or misreporting that is correlated with the state space for certain estimation steps. For example, estimating conditional exit probabilities clearly suffers from bias if measured exit probabilities are biased over the state space. Two potential issues may be particularly important in our case: non-reporting due to the cutoff of 20 workers and that our measure of exit may inflate exit towards the end of our data because we cannot distinguish permanent exit from temporary non-reporting. We think that both issues likely introduce biases that are small in magnitude.

As for the cutoff of 20 workers, the issue would be particularly problematic if plants regularly moved back and forth over the threshold or if plants with more than 20 workers moved permanently below the threshold, which we would wrongly classify as plant exit. We do not think that these are important issues in the Indonesian data. For example, few plants shrink and as we show in the main text, plants with 20 workers become relatively less important vis-a-vis larger plants over time. Also, given that pre-1990, plants often continue reporting even if they move below 20 workers, we find that movements around the threshold are rare. As for the classification of exit towards the end of the sample, we note that non-reporting actually seems to decline over time and 2015 (the last year of the data) is a census year in which enumeration is most complete.

2.A.2 Further main descriptives

Figure A.1 reports the evolution of the share of employment and value-added output that is captured by the Indonesian manufacturing plant census (1975-2015) in comparison to aggregate manufacturing value-added output and employment as reported in the GGDC 10-sector

Figure A.1: Representativeness of manufacturing panel over time

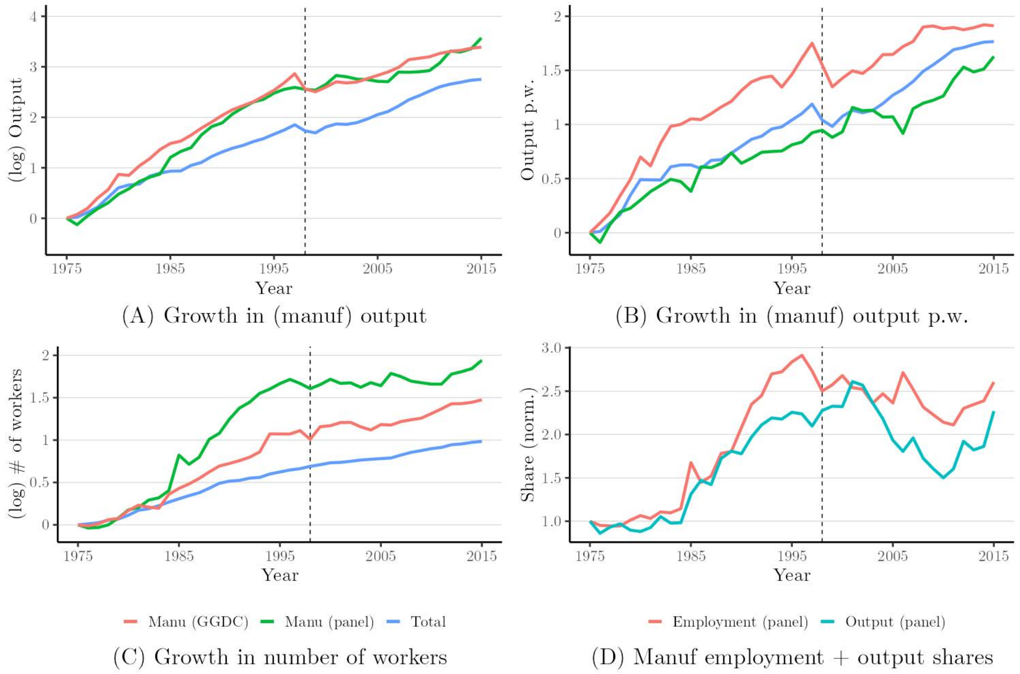
Notes: Evolution of the employment and value-added output share captured by the Indonesian manufacturing plant census (1975-2015) in comparison to aggregate manufacturing value-added output and employment as reported in the GGDC 10-sector database (1975-2012) and the Economic Transformation Database (1990-2018). Straight lines report (unweighted) time averages. The black dotted line highlights the Asian Financial Crisis.

database (1975-2012) and the Economic Transformation Database (1990-2018).[25] The series tend to increase over time, which captures the fact that medium- to large-size manufacturing plants become more important over time. However, there is important variation in the shares over time that is not well-captured by a simple secular increase in the importance of medium- to large-sized manufacturing plants. In the model, we allow for the time-varying importance of the rest-of-the-economy, and differentially so for output and employment, including the part of manufacturing that our data misses. Note that for output, the increase is much stronger if we do not clean the value-added series, because a few plant entries can have an outsized effect on total value-added (e.g. for only a minimally cleaned output series, we found that the output share of our manufacturing data can increase to up to 80% by 2015, entirely driven by a few plant entries).

In the main text, we also report that our manufacturing panel misses 99% of manufacturing plants in Indonesia. This is based on information on a random five percent sample of all manufacturing establishments from the Indonesian Economic Census in 2006 reported in Hsieh and Olken (2014). We also verify that our micro-data is consistent with capturing all manufacturing plants with more than 20 workers. For example, based on the 2006 census sample (as reported in Hsieh and Olken (2014)), manufacturing plants with more than 50 workers should capture 34% of total manufacturing employment, while this figure is 32% based on employment in our micro-data (29.5% after cleaning) and taking the aggregate sectoral employment from the GGDC 10-sector database as denominator. Given that the manufacturing plant panel includes new plants based on the Economic Census, coverage is more complete after Economic Census years.

[25]We merge the latter two series consistently over time by enforcing the more recent vintage and aligning all series before 1990 to be consistent with the evolution after 1990. Specifically, for each variable, we take the ratio of the GGDC10 and ETD series in 1990 to be the amount that the 1990 GGDC series needs to be adjusted by. We then similarly correct each year from 1975 to 1990 by a correction factor that equals unity in 1975 (no correction in 1975), is equal to the full correction in 1990 and is taken from an equal-spaced, smoothed series in the years between.

Figure A.2: Evolution of aggregate and sectoral employment and output (in logs)



Notes: (Economy-wide) Total and GGDC are based on joining the GGDC 10-sector Database (1975-2012) and the Economic Transformation Database (1990-2018). Panel refers to the Indonesian manufacturing plant census (1975-2015, 20+ workers). All series are normalized by their respective value in the first year. (A) and (B) use value-added output.

Next, Figure A.2 reports the evolution of aggregates in the Indonesian economy, showing the series in logs to better visualize how growth rates changed over time.

For completeness, Figure A.3 reports the full year-to-year evolution of employment shares for different plant sizes.

We further report estimated Pareto tail coefficients for the manufacturing data in Figure A.4. We follow Chen (2022) in constructing two simple, but alternative measures of the Pareto tail. Let $F(x,t)$ be the CDF of the underlying distribution of plant employment, and f the density function. Then $\tilde{F}(x,t) \equiv 1 - F(x,t)$ denotes the fraction of plants with size greater than x, and $\tilde{F}^{emp}(x,t) \equiv \int_x^\infty y dF(y,t)$ the total employment in plants with size greater than x. In addition, let T_L be the employment size threshold for large plants and T_S for small plants. Assuming that $F(x,t)$ follows a Pareto distribution with shape parameter k_t, we have:

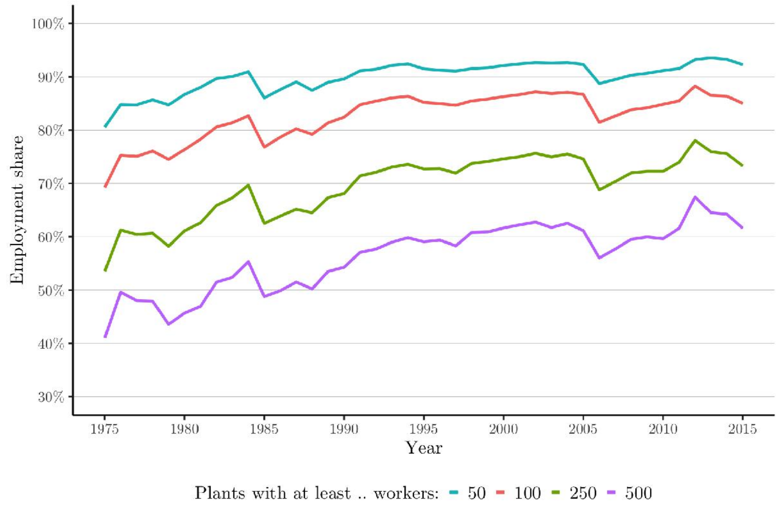
$$k_t = 1 - log\frac{\tilde{F}^{emp}(T_L)}{\tilde{F}^{emp}(T_S)}/log\frac{T_L}{T_S}$$

Alternatively,

$$k_t = -log\frac{\tilde{F}(T_L)}{\tilde{F}(T_S)}/log\frac{T_L}{T_S}$$

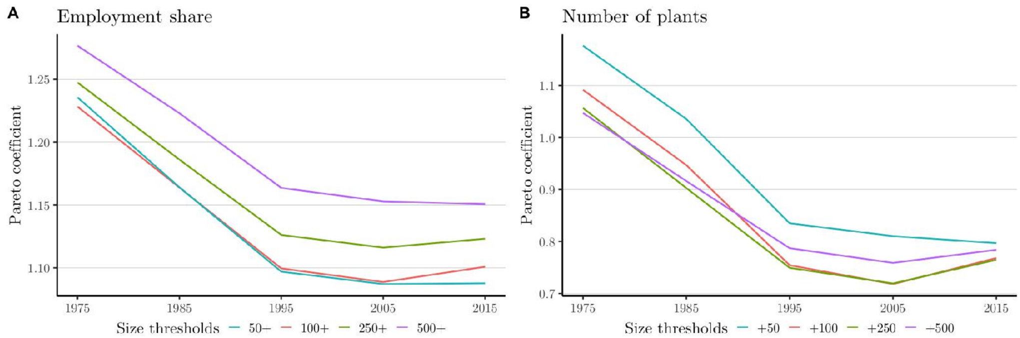
In both cases, Pareto tails can be computed in the absence of knowing the employment share or fraction of plants below 20 workers, because these shares cancel out. Panel A reports estimated Pareto coefficients for different thresholds T_L for the employment share measure, while Panel B reports the estimated Pareto coefficients for the same thresholds T_L but for the number of plants instead.

Figure A.3: Evolution of employment shares in large Indonesian manufacturing plants



Notes: Data based on Indonesian manufacturing plant census (1975-2015) with 20+ workers.

Figure A.4: Evolution of Pareto tail coefficients



Notes: Data based on Indonesian manufacturing plant census (1975-2015) with 20+ workers.

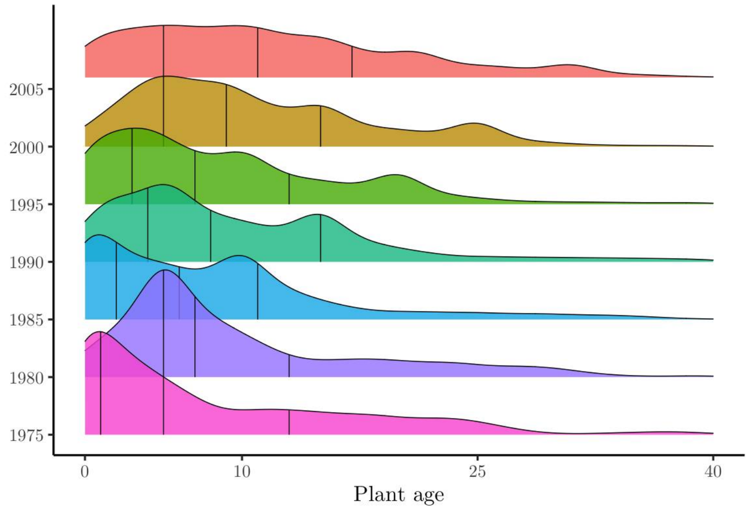
In accordance with the main text, we estimate Pareto tails by decade in 1975, 1985, 1995, 2005 and 2015. The overall trend in Pareto tails is consistent for different measures and different thresholds T_L: the tail of the employment distribution grows markedly thicker over time. However, there are important differences both across measures and across different thresholds, which is not in line with a common Pareto distribution in the cross-section. The quantitative implications are also very different for the two different measures, because Pareto tails below 1 imply that not even the mean of the distribution is defined.

Figure A.5 reports changes in the plant age distribution over time. Average plant age increased by roughly 40% between 1975 and 2006. While the 1975 plant distribution does feature very old plants, by far most plants are very young. In contrast, 30 years later, the plant age distribution is far more equally distributed, featuring more medium-old plants and relatively fewer very young plants.

2.A.3 Additional results for iterating on initial distribution

In this subsection, we provide further results on the reduced-form exercise of iterating on the discretized initial plant distribution. We start out by showing that 1975 and 1976 are good starting years, and if anything, give conservative estimates. We then show that results are very

Figure A.5: Evolution of the age distribution across plants



Notes: Data based on Indonesian manufacturing plant census (1975-2015) with 20+ workers. Only showing data for years 1975, 1985, 1995,2006. The last year is 2006, because 2006 is the last year where plant age is separately asked in the survey. After 2006, we only observe plant age for surviving plants, biasing estimates of the cross-sectional age distribution.

similar when accounting for entry and exit.

Results are similar when varying transition matrices and averaging transition matrices over multiple years. From an historical point of view, taking 1975 as the starting year and 1975 and 1976 as the initial years from which we construct the reduced-form transition matrix, is conservative, because (1) the two years fall in between the two periods of growth accelerations identified by Hausmann, Pritchett, and Rodrik (2005) for Indonesia (which are dated to 1967-1974 and 1987-1994 respectively), and (2) they are not affected by any notable labor or financial market reforms and predate the major tax reform of 1976 (see: Hill 2000). This is not to say that the 1970s were economically without important events. Oil prices rose dramatically in 1973 and inflation became a major macroeconomic issue that was followed by interest rate hikes and ceilings on commercial bank credit in 1974 (Hill 2000, 294). There were also important export-promoting trade policy reforms throughout the 1970s, but during a time in which Indonesia was still a very closed economy. Based on World Bank national accounts data, exports made up around 22% of GDP in 1975 whose share actually slightly decreased from 1975 to 1976, alleviating the concern that the growth between 1975 and 1976 is purely driven by trade reforms.

Figure A.6 shows that taking transition matrices for any other starting year (e.g. 1985 as starting year for 1985-1986 transitions) gives, if anything, stronger results than the ones reported for 1975 and 1976. Most years see much more growth in the average plant size and the employment share of large plants. Importantly, all years show an eventual increase in the employment share of large plants, giving credence to the idea of a tail that slowly fills up. Furthermore, any other starting year in the 1970s would have given much stronger results. E.g. taking transitions between 1976 and 1977 would have explained 67% of the average size increase and 96% of the employment share increase over time. We also considered averaging transition matrices across multiple years and obtained very similar results.

Next, we considered two variations on the exercise to account for entry and exit. To begin with,

Figure A.6: Reduced-form transition dynamics from initial conditions in 1975 and all year-to-year transition matrices

Notes: Reduced-form transition dynamics implied by initial plant size distribution in 1975, but taking transition matrices from each year-to-year pair in the data.

note that entry and exit is potentially very important, especially if entering plants differ from exiting plants. Of the roughly 6,800 plants with more than 20 workers operating in 1975, less than 12% were still operating in 2015. On the other hand, as shown in Figure 2, the number of active plants increased by a factor of 4 between 1975 and 2015. This means that the vast majority of active plants in 2015 did either not exist or was not captured in the 1975 census. To capture the role of entry and exit, we amend the previous exercise by including a state-0 which captures inactive plants or potential entrants. This means that both the initial distribution is defined over an additional state-0 and the transition matrix will feature transitions into (exit) and out of state-0 (entry). To construct the new transition matrix, we can use observed entry and exit flows. Since transition matrix entries are computed as the share of flows from bin x in period t into any other bin in period $t+1$, we can readily compute transitions from an active state to an exit state. However, we cannot directly compute entries from inactivity, because the baseline is fundamentally undetermined. We do not know how many inactive or potential plants there are. This means we can also not directly compute the new initial distribution that includes the measure of plants in state-0. Since both the transition matrix and the initial distribution depend on the number of inactive plants, this number cannot be identified from observables in the first two periods alone. In theory, we can pin down the initial number of inactive plants by enforcing that the transition matrix stays constant over time and by feeding in another moment, the change in the number of plants between 1976 and 1977. However, the initial periods saw an initial decrease in the number of plants between 1975-1976 and a subsequent increase between 1976-1977. To match this pattern, we would have to enforce a negative transition matrix entry for staying inactive.

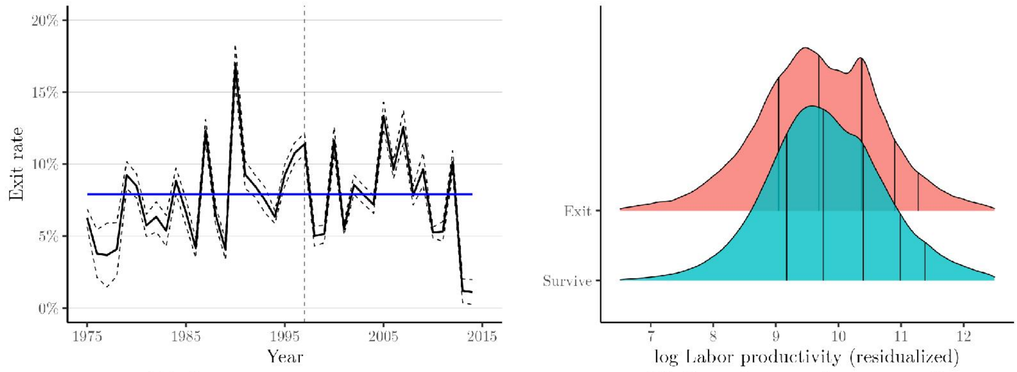
To avoid this, while giving almost indistinguishable results, we instead assume that the share of inactive plants that stay inactive is 0. This identifies the transition matrix and we then consider two additional exercises where we keep this transition matrix fixed. In the first version of the exercise with entry and exit, we simply iterate on the initial distribution and the transition matrix. This keeps the total number of plants (inactive + active) constant, while introducing interesting entry and exit dynamics that directly affect the evolution of the plant size distribution over time. Results for this exercise are given by the lines "Entry + Exit" in Figure A.7. While the long-run results are almost unchanged to the previous results, introducing entry and exit does speed up transition dynamics considerably, providing a much better out-of-sample fit for the early transition period. This is driven by observed exiting plants being smaller and less productive than observed new entrants in 1976. With a positive share of inactive plants staying inactive each period would slow these predicted transition dynamics down.

In the second version of the exercise with entry and exit, we additionally vary the number of plants that enter each period. Specifically, we exactly match the increase in the number of active plants over time as shown in Figure 2.2, while taking information on new entrants and exits only from 1975 and 1976. In contrast to the previous exercise with entry and exit, here we do take limited information on future plant entry and thus it does not lend as well to

Figure A.7: Reduced-form transition exercise with entry and exit

Notes:

Figure A.8: Main plant exit patterns



(A) Avg exit rates over time (B) Distribution: Survive vs Exit

Notes: Panel A: Residualized by 5-digit industry fixed effects. Standard errors are two-way clustered by industry and year. Results show 95% confidence bands and blue line gives unconditional average. Panel B: Labor productivity is measured as value added per worker, residualized by 5-digit industry-year fixed effects. Vertical lines report 25th, 50th, 75th, 90th and 95th percentiles respectively.

predicting future changes in the plant size distribution. However, this exercise gives a more complete picture of the importance of entry and exit observed in the data. Results are given by the lines denoted "Entry + Exit Growth" in Figure A.7. The series again behave very similarly as before, but we can more clearly see that important year-to-year fluctuations in the real data series are driven by entry shocks. For example, the inclusion of many more plants in 1985 had important medium- to longer-run effects on the evolution of the size distribution.

2.A.4 Further details on exit behavior

Here, we provide evidence that exit behavior only varies little with plant productivity and does not clearly respond to aggregate shocks. Figure A.8 Panel A shows that exit rates vary quite strongly over time, but are not straightforwardly affected by measurable aggregate economic shocks. For example, exit rates actually decreased during the Asian Financial Crisis in 1998 & 1999. To focus only on within-industry variation, we residualize exit rates by 5-digit industry fixed effects here. Figure 5 Panel B shows productivity distributions of exiting and surviving plants using value added per worker as a simple measure of (labor) productivity and only using within-industry-time variation by residualizing the measure by detailed 5-digit industry-year fixed effects. Surviving plants are more productive on average than exiting plants, but given strong overlap in the two distributions, most plants do not exit because of their productivity. A dynamic implication of this difference is that much of plant exit is not driven by productivity so that it takes time for unproductive plants to leave the economy and productivity improvements from selective exit take time to materialize.

2.A.5 Details and robustness for production function estimation

We start out by proving formal identification of the production function for the different cases (static vs. dynamic capital, time variation in production functions, industry heterogeneity). We then discuss our estimation strategy and provide detailed estimation results for the case with full flexibility on the time variation in production functions but without industry heterogeneity. In the last part, we then consider the case of industry heterogeneity.

Identification of production function

We have the following setup. Log output by firm i at time t is given by

$$y_{it} = x_{it} + f(h_{it}, k_{it})$$

where x_{it} is productivity, h_{it} is labor input, k_{it} is capital and the price of the homogeneous production good is normalized to unity throughout. We leave $f()$ unspecified here to make clear that the identification proof is non-parametric. In the estimation and model, we assume that $f()$ is Cobb-Douglas. We also suppress industry variation here, but identification extends naturally to the case with industry variation in production functions. We start with the more general case where both capital and labor are chosen dynamically and then discuss the simpler case when capital is statically chosen.

Following the literature, we assume that in the case of dynamic capital input choices, capital is pre-determined. The input choices can then be written as non-parametric functions of the relevant state-space:

$$\begin{aligned} h_{it} &= f_h(h_{it-1}, k_{it}, x_{it}, \Omega_t) \\ k_{it} &= f_k(k_{it-1}, h_{it-1}, x_{it-1}, \Omega_{t-1}) \end{aligned}$$

where we have specified in which sense capital is pre-determined. For notational simplicity, we drop the dependence on Ω_t throughout, because identification arguments are cross-sectional. At last, productivity follows a general first-order Markov process with

$$x_{it} = f_x(x_{it-1}, u_{it}) \qquad with \quad u_{it}|x_{it-1} \sim U(0,1)$$

where u_{it} is an innovation. This representation follows the Skorohod representation of random variables and is without loss of generality (see Demirer 2022). Output, labor and capital are strictly monotonic in productivity, which imposes weak regularity conditions on the productivity process x_{it}, such that each can be inverted for productivity:

$$\frac{\partial y}{\partial x} > 0 \implies x_{it} = f_y^{-1}(h, k, y, \Omega_t) \tag{2.15}$$

$$\frac{\partial h}{\partial x} > 0 \implies x_{it} = f_h^{-1}(h_{-1}, h, k, \Omega_t) \tag{2.16}$$

$$\frac{\partial k}{\partial x} > 0 \implies x_{it-1} = f_k^{-1}(k, k_{-1}, h_{-1}, \Omega_{t-1}) \tag{2.17}$$

We now adapt the identification proof by Demirer (2020):

$$\begin{aligned} &h = f_h(h_{-1}, k, f_x(x_{-1}, u)) = f_h(h_{-1}, k, f_x(f_y^{-1}(h_{-1}, k_{-1}, y_{-1}), u)) = \tilde{f}_h(h_{-1}, k, k_{-1}, y_{-1}, u) \\ &u = F_{h|h_{-1},k,k_{-1},y_{-1}}(h|h_{-1}, k, k_{-1}, y_{-1}) \end{aligned}$$

Intuitively, two firms with the same current capital, previous labor, previous capital and previous output, but different today's labor differ only in innovation to productivity. Using the identified u, we can then identify the production function using the control function $f_x(x_{-1}, u)$ for unobserved productivity x:

$$y = f(h, k) + f_x(x_{-1}, u) = f(h, k) + f_x(f_y^{-1}(h_{-1}, k_{-1}, y_{-1}), u)$$

A semi-parametric regression of y on the known function $f(h, k)$ of observables and a non-parametric term in observables/identified terms $(h_{-1}, k_{-1}, y_{-1}, u)$ identifies the output elasticities of interest.

In the case where capital is chosen statically (e.g. via a frictionless rental market), the identification approach simplifies. Specifically, dependence on k drops out in the sense that input choices are now given by:

$$\begin{aligned} h_{it} &= f_h(h_{it-1}, x_{it}, \Omega_t) \\ k_{it} &= f_k(h_{it-1}, x_{it}, \Omega_t) = f_k(h_{it}, x_{it}, \Omega_t) \end{aligned}$$

Output and labor are strictly monotonic in productivity such that:

$$\frac{\partial y}{\partial x} > 0 \implies x_{it} = f_y^{-1}(h, y, \Omega_t) \tag{2.18}$$

$$\frac{\partial h}{\partial x} > 0 \implies x_{it} = f_h^{-1}(h_{-1}, h, \Omega_t) \tag{2.19}$$

Identification is then given by:

$$\begin{aligned} h &= f_h(h_{-1}, f_x(x_{-1}, u)) = f_h(h_{-1}, f_x(f_y^{-1}(h_{-1}, y_{-1}), u)) = \tilde{f}_h(h_{-1}, y_{-1}, u) \\ u &= F_{h|h_{-1}, y_{-1}}(h|h_{-1}, y_{-1}) \end{aligned}$$

Now, without dependence on capital, two firms with the same previous labor and previous output, but different today's labor differ only in innovation to productivity. Using the identified u, we can then identify the production function using the control function $f_x(x_{-1}, u)$ for unobserved productivity x:

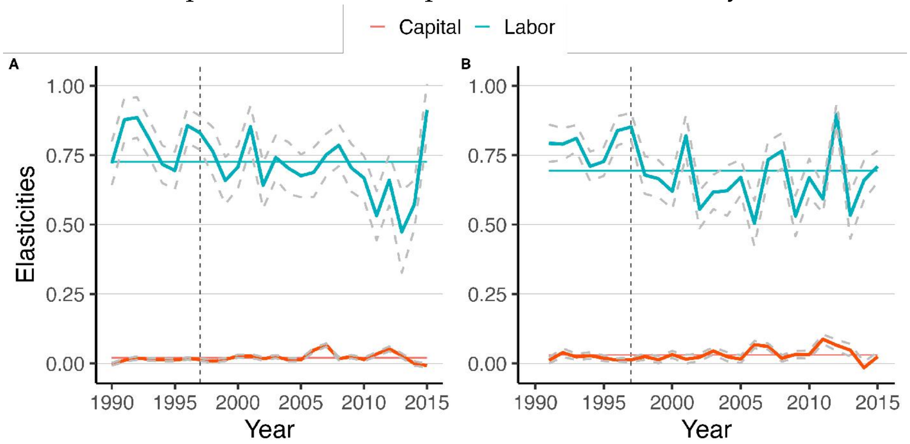
$$y = f(h, k) + f_x(x_{-1}, u) = f(h, k) + f_x(f_y^{-1}(h_{-1}, y_{-1}), u)$$

A semi-parametric regression of y on the known function $f(h, k)$ of observables and a non-parametric term in observables/identified terms (h_{-1}, y_{-1}, u) identifies the output elasticities of interest.

Production function estimation with time-variation but no industry variation

In the following, we report our production function estimation results. We start by showing results based on the sample from 1990 until 2015, which includes data on plant-level capital. We then discuss estimates from 1975-1990 and estimates with further industry heterogeneity. Importantly, (1) the estimated elasticities for labor are not biased by excluding capital, and (2) the estimated capital elasticities post-1990 are very low, meaning that any choice on how to model capital has only very small effects on productivity estimates.

Figure A.9: Estimated capital and labor output elasticities for each year between 1990-2015



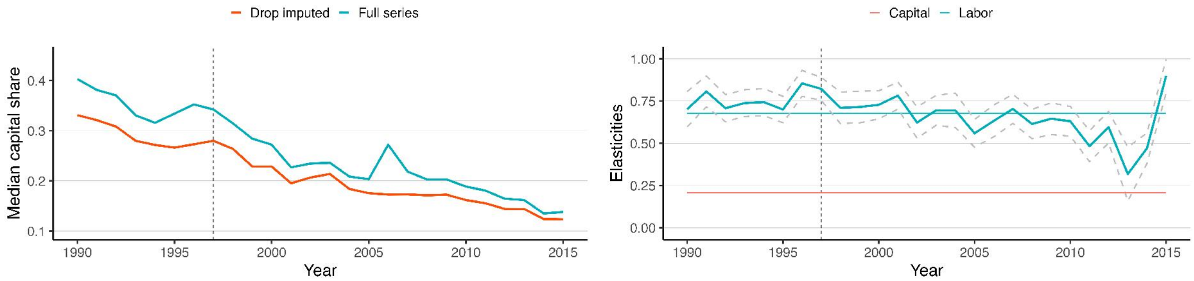
Notes: Panel A gives estimates based on a static choice of capital, Panel B gives estimates assuming capital is dynamically chosen. The latter requires to drop year 1990 in the estimation, because the estimator requires previous capital choices. Horizontal lines give average estimates over time. Grey dotted lines give 95 percent confidence bands (note that standard errors are not yet corrected for the two-stage estimation).

Figure A.9 reports estimated capital and labor elasticities. Panel A gives estimates based on assuming that capital is statically chosen, while Panel B allows capital to be dynamically chosen. The estimates assume a common production function across manufacturing industries, but allow for fully flexible elasticities over time. Allowing for time-series variation is important, because policy functions are generally time-varying if the economic environment changes so that pooling estimates across years without allowing input choices to vary over time is model inconsistent. Elasticity estimates are remarkably stable and do not show a clear trend over time. Estimates are also very similar whether one assumes static or dynamic capital input choices. For example, the average estimated labor elasticity varies by less than 5% across the two different estimators (from 0.726 to 0.694). Furthermore, the estimated labor elasticity is close to 2/3, a common value in the literature. Note, however, that this is in a context where both the aggregate labor share in manufacturing (around 0.25) and the median labor share (around 0.54) are substantially below the estimated elasticities. Our model accounts for this systematic difference. Estimated capital elasticities are much lower than commonly estimated/used values in the literature and we discuss this point further below.

Elasticity estimates for each year are based on the following estimation steps: In the first step, we flexibly estimate the rank (taking the empirical cumulative distribution function) of labor conditional on previous labor and previous output. From this estimate, we then back out a monotonic transformation of the productivity innovation u as the difference between the observed and estimated rank. In the second step, we then estimate a log-log regression in capital and labor on output, but flexibly controlling for the different components of the control function (the estimated u and previous labor, output and capital). The dynamic and static capital estimators differ only in the variables that we condition on in each of the two estimation steps. For both estimators and both estimation steps, we draw on generalized additive models (GAMs) as a flexible and robust way to estimate semi-parametric models (Hastie 2017). We obtain very similar results when choosing flexible polynomial regressions.

Given our estimation approach, why are capital elasticity estimates so low? We think the main reason here is a standard attenuation bias in the capital elasticity estimates given substantial

Figure A.10: The effects of alternative estimates of capital elasticities



Notes: Left: Evolution of capital shares based on the capital series in Cali, Le Moglie, and Presidente (2021). For better comparability, using also their real value added series which deflates output by sector-specific prices. To construct capital shares, assume that rental rate is 14 percent (interest rate of 4 percent and depreciation of 10 percent) and value added tax is 10 percent. Right: Estimated labor elasticities when assuming that capital is statically chosen and its output elasticity is fixed at the median from Cali, Le Moglie, and Presidente (2021), which is 0.207.

measurement error in observed plant-level capital. Apart from the control function term, the second step estimation is a standard linear regression in capital, so that any classical measurement error in capital will attenuate the estimated capital elasticity. Why do we suspect measurement error in the capital series? As reported in Cali, Le Moglie, and Presidente (2021), one common issue in the reported plant-level capital series is misreporting in the units, which exactly shows up as a log-additive measurement error. Apart from such unit misreporting, capital – as is well known – is also more susceptible to misreporting because it is a stock that not all plants necessarily keep track of (in contrast to cost flows such as the labor bill). Inferring changes in the stock directly from reported investments and assumptions on capital-type-specific depreciation rates (as is done in perpetual inventory methods and the capital series based on Cali, Le Moglie, and Presidente (2021) that we draw on), can mitigate some of this measurement error but unlikely all.

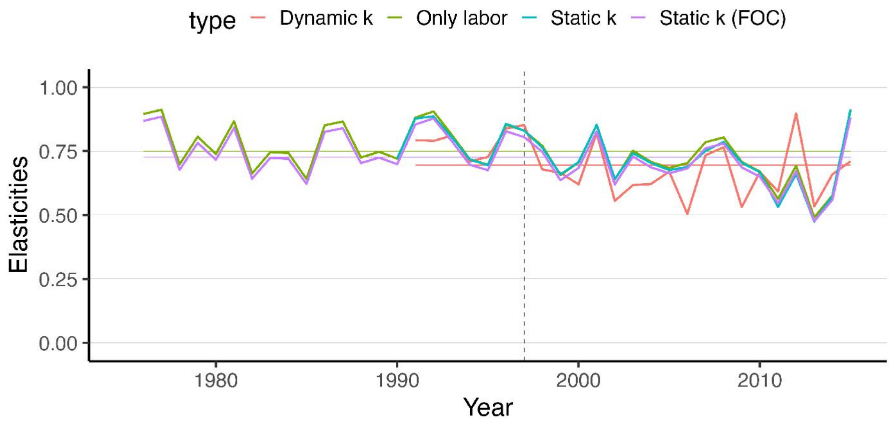
Besides attenuation bias from noisy capital reporting, the capital series likely also suffers from more systematic biases that complicate their use. To show this, we consider the case where capital is chosen statically in which case we can directly make use of plants' first-order conditions instead of estimating the capital elasticity from the output regression. This first-order approach does not generally suffer from attenuation bias because one can estimate capital elasticities from average or median capital shares, which are robust to log-additive measurement error. Specifically, the static capital input choice conditional on the assumed production structure implies the standard condition:

$$\alpha_{jt} = \frac{(r_t + \delta_t)k_{it}}{(1 - \tau_y)p_t y_{it}}$$

where α_{jt} gives the potentially industry-time-specific capital elasticity, $(r_t + \delta)$ gives the competitive rental rate of capital, τ_y gives a value-added tax and $p_t y_{it}$ gives plant revenue.[26] Figure A.10 (left) plots changes in median capital shares over time. Capital shares more than halved between 1990 and 2015, which would also imply a halving of capital elasticities in the case of static capital choices. This seems unlikely, among others because of global capital-biased technological change, the advent of industrial robots and a running out of labor intensive industrialization in Indonesia. Also, in the case of adjustment frictions, one would have expected

[26]If observed capital features log-additive measurement error ($\tilde{k}_{it} = k\varepsilon$), then, $\mathbb{E}_i \frac{(r_t+\delta_t)\tilde{k}_{it}}{(1-\tau_y)p_t y_{it}} = \alpha_{jt}$.

Figure A.11: Estimated output elasticities of labor for each year between 1975-2015



Notes: Dynamic and static k report previously estimated labor elasticities. Only labor shows estimates based on the assumption that the production function only features labor. Static k (FOC) assumes that the capital choice is static and enforces the average estimated capital elasticity from the Dynamic k estimator (which nests the static k estimator) from 1990-2015.

an increase in the capital share as the economy is catching up, not a decline. More likely, we think plants systematically underreport new capital investments and we might overestimate depreciation of existing capital.[27] Taking noisy capital estimates together with systematic misreporting, it is hard to trust the Indonesian plant-level capital series. We thus choose to stick with the low estimated capital elasticities for the baseline model and results, which implies that observed capital variation has very small effects on plant output and labor decisions. Still, to gauge how sensitive our estimates are to low capital elasticities, Figure A.10 (right) also reports estimated labor elasticities in case where we enforce the much higher capital elasticity based on the median capital share. We find that estimated labor elasticities are almost unchanged when assuming such a higher capital elasticity.

Next, we consider production function estimation for the period 1975-1990 for which we lack data on capital. We follow two different approaches to understand whether there have been important changes in production functions over time. In the first approach, we simply assume that the production function does not feature capital and estimate only labor elasticities from 1975-2015, comparing this to the estimated labor elasticities estimated with capital data from 1990-2015. In the second approach, we assume a static choice for capital and enforce the estimated capital elasticities from before. As shown in Figure A.10, we find that for both approaches estimated labor elasticities are very similar to estimates after 1990 and that they are remarkably stable and do not show a clear trend over time. We interpret this as strong evidence that production functions did not systematically change over time.

Production function estimation with industry variation

We now consider industry-level variation in production functions. (Show 2 results: 1. Test for equality across industries. 2. Check how correlated productivity estimates are)

[27] Of course, one may also explain a strong decline in estimated revenue elasticities by changes in markups. However, the required magnitude of such markup changes also seems unrealistic. Through the lens of a standard monopolistic competition model, the difference in estimated revenue elasticities would translate to a roughly 100 percentage point increase in the markup. That is, if a product is sold at 50% above marginal cost, it would now be sold at 150% above marginal cost.

2.A.6 Details and robustness for aggregate technology estimates

In this section, we give a formal identification proof, estimation details for separating aggregate technology from idiosyncratic productivity and a discussion of the drivers of technology growth. To simplify the exposition, in the following we will denote the logarithm of a vector by lower letter cases. Hence, in slight deviation from the exposition in the main paper, we assume that productivity for plant i at time t is given by: $Y_{it} = Z_t exp(s_{it})$. So that the log-additive form is: $y_{it} = z_t + s_{it}(s_{it-1})$. The average of within-plant changes in log productivity is then:

$$\frac{1}{N^S_{t,t-1}} \sum_{i \in \mathcal{N}^S_{t,t-1}} \Delta y_{it} = \underbrace{z_t - z_{t-1}}_{\Delta z} + \underbrace{\frac{1}{N^S_{t,t-1}} \sum_{i \in \mathcal{N}^S_{t,t-1}} \Delta s_{it}}_{\text{Avg mean reversion of survivors}}$$

Identification

Proposition 5 (Main identification result)**.** *Under the following four assumptions:*

1. *(**Common first-order stationary Markov process**) s follows the same general first-order, stationary ergodic Markov process for all i & t.*
2. *(**Selective exit**). The decision to exit after period t can flexibly depend on observables and unobservables X_{it} as well as productivity s_{it}, but may not depend on future productivity s_{it+1}. Specifically,*
$$\mathbb{P}(exit) = f(X_{it}, s_{it}, z_t) \quad with \quad \mathbb{P}_t(exit) \perp\!\!\!\perp s_{i,t+1} | s_{i,t}$$
3. *(**No complete exit over s**) $\mathbb{P}_t(exit|s_{it}) < 1 \, \forall s \in Supp(s)$*
4. *(**Connected support in s**) For each period t, there exists at least a subset of the support of s in that period which is fully contained in the support of all s in all future periods. Formally: $\forall t, \exists S_t \subset Supp(s_{it})$ for which $S_t \subset \cup_{\tau > t} Supp(s_{i\tau})$.*

the path $z_t \, \forall t$ is identified given some normalization z_τ for some $\tau \in [0, T]$ and $\max t \equiv T \to \infty$.

Proof. To already convey the idea of a suitable estimator for the time path of z_t, let us proof Proposition 1 constructively. Identification proceeds sequentially in two fundamental steps. In the first step, I show identification of the density of the stationary distribution of s, which is identified for $t \to \infty$. In the second step, the density of the stationary distribution is used to identify the path of z_t backwards by starting at some final time T. The density of the stationary distribution is key because it can be used to construct weights under which a weighted difference Δy_{it} exactly identifies Δz_t. Specifically, there exist weights ω_s such that $\sum_{i \in \mathcal{N}^S_{T+1,T}} \omega(s_{iT})(s_{iT+1} - s_{iT}) = 0$ (where $\sum_i \omega_s(s_i) = 1$). These weights recover the stationary distribution of s. Denote by $f^{SS}(s)$ the density of the stationary distribution at s and by $f_t(s)$ the density of the distribution of s at time t. Assuming that this distribution shares the support of the stationary distribution, we have:

$$\lim_{N \to \infty} \sum_{i \in \mathcal{N}^S_{t+1,t}}^{N} \frac{f^{SS}(s_{it})}{f_t(s_{it})} \Big(log(s_{it+1}) - log(s_{it}) \Big) = 0$$

The weights are thus defined by $\omega_s(s_{it}) \equiv \frac{f^{SS}(s_{it})}{f_t(s_{it})}$ and are a function of the unknown density function of the stationary distribution of s. To identify the density $f^{SS}(s)$, start with the distribution of plants at t_0 over known y_{i0}. The idea is to follow survivors (as they follow the process for s), while replacing exiting plants with plants that stay in the panel that have similar y_{it}. More formally, denote the initial set of plants by $\mathcal{N}_0$ where each plant is given a uniform weight $\tilde{\omega}_{i0} = \frac{1}{N_0}$. We are interested in updating $\mathcal{N}$. For this, pass on the weight of each surviving plant and redistribute the weight of each plant that exits to close plants around them.[28] This gives $\mathcal{N}_1$. Updating in this way allows to eventually pass on weight to plants that have entered the economy, even if they have entered in an arbitrarily selective way. As $t \to \infty$, surviving plants will eventually populate the entire support of s and this procedure gives a synthetic sample $\mathcal{N}_\infty$ with weights $\tilde{\omega}_{i\infty}(s_{i\infty})$ that directly identify the density $f^{SS}(s)$.

The second step of the proof takes the identified density $f^{SS}(s)$ and works backwards from time T. Normalizing the final value z_T, one can show that z_{T-1} solves a fixed point problem. Specifically:

$$\sum_{i \in \mathcal{N}^S_{T,T-1}} \omega_{\hat{s}_{T-1}(z_{T-1})}(y_{iT} - y_{iT-1}) = z_T - z_{T-1} + \sum_{i \in \mathcal{N}^S_{T,T-1}} \omega_{\hat{s}_{T-1}(z_{T-1})}(s_{iT} - \hat{s}_{iT-1}(z_{T-1})) = z_T - z_{T-1}$$

where the last equality holds only if the guess z_{T-1} is correct. It thus gives a nonlinear equation in z_{T-1} (since the weights and the right-hand side depend on z_{T-1}). One can iterate on this procedure to identify the path of z_t backwards. At any point in time $t < T-1$, one can also alternatively guess z_{T-1} and instead of using weights at all, estimate the bias term $\sum_{i \in \mathcal{N}^S_{T,T-1}}(s_{iT} - \hat{s}_{iT-1}(z_{T-1}))$ directly using future survivors with similar s. This alternative relaxes the assumption of a common support with the stationary distribution and instead only requires that we can build a sample with similar survivors – requiring a much weaker connected support.

Estimation

Estimation proceeds along the lines of the constructive identification proof. In the first step, one sequentially builds the synthetic panel with weights $\omega_s(s_{it})$ (which sum to 1 in each year). In principle, one can use any standard matching estimator for passing on the weight for exiting plants. We find that a Kernel matching estimator works well, because matching is only based on one variable and the Kernel estimator distributes the weight widely across multiple observations, reducing variance.[29]

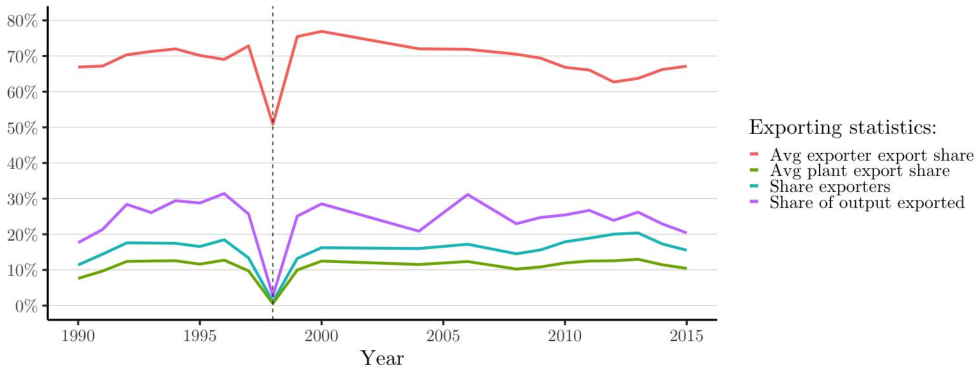
One can then estimate $f^{SS}(s)$ using observed s in the last period T and constructed weights $\hat{\omega}_s(s_{iT})$. Any standard density estimator such as a Kernel density estimator works here. To reduce variance, one can also estimate $f^{SS}(s)$ on the last x periods (where x is at the discretion of the researcher). In general, for any fixed T, the bias on the estimated weights is increasing in the persistence of the process as well as in the distance of the initial distribution from the stationary distribution. That is, for large T and low persistence, one can use more periods in the end to estimate $f^{SS}(s)$.[30] Once the density is estimated, one can then proceed by sequentially

[28] As $N \to \infty$ and the assumption that exiting probabilities are always strictly lower than one, there always exists a plant that is arbitrarily close to an exiting plant.

[29] Note that one can readily match based on further variables such as detailed industries to minimize the risk of model misspecification.

[30] A formal treatment of optimally solving this trade-off is beyond the scope of this paper.

Figure A.12: Evolution of key exporter statistics



Notes: Data based on Indonesian manufacturing plant census (1975-2015) with 20+ workers and reported export shares (out of total value-added).

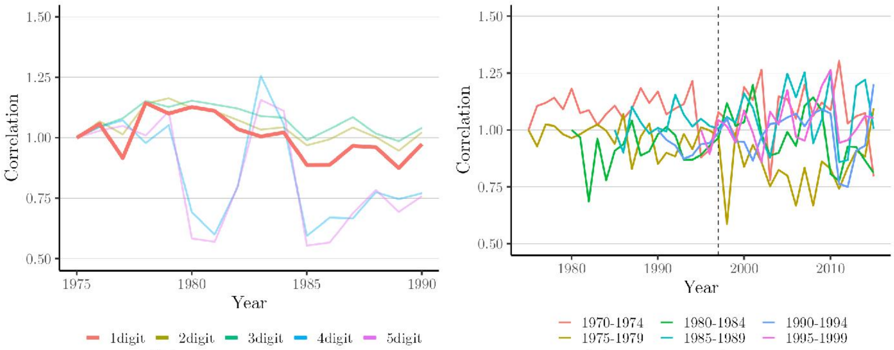
estimating the path z_t. For each period t and for each guess of z_t, this means one has to estimate $f_t(s_{it}(\hat{z}_t))$. Again, any standard density estimator works here. One can then construct the weights according to: $\omega_{st}(s_{it}) \equiv \frac{f^{SS}(s_{it})}{f_t(s_{it})}$. Alternatively, one can choose not to use weights and instead directly estimate the bias from mean reversion. In that case, one can again use any kind of matching estimator to match plants in t with productivity $s(\hat{z}_t)$ to future survivors with similar s. The variance in the bias estimate reduces with the number of matched plants such that one to many matches are recommended. As before, a Kernel-based matching estimator is a natural choice here. In either the approach with weights or with an estimated bias term, one then finds z_t that solves the fixed point problem, requiring a standard root finder. We have not formally proven uniqueness of the root, but in practice, we found no issue of a multiplicity of roots. In principle, any (weighted) moment of within-plant changes in productivity that preserves scalar multiplicity can be used for the estimation. In practice, we use (weighted) median changes in productivity as the median is less susceptible to outliers. Results are similar when taking the weighted average.

The drivers of aggregate technology

We now discuss the potential drivers of the estimated aggregate technology growth path. First, through the lens of standard endogeneous growth models such as Romer (1990), the increase in technology growth after the year 2000 could be driven by overall human capital improvements and a better integration into global markets. However, there seems to be no sharp change in human capital improvements nor in the integration into global markets. For example, Figure A.12 shows that the share of manufacturing output that is exported stays very stable around 20% and the fraction of plants that are exporting also remained flat since the mid 1980s.

Alternatively, the patterns may be in line with models of learning and imitation (e.g. Perla, Tonetti, and Waugh 2021), whereby the initial entry of many new and relatively unproductive plants lowered productivity growth and the subsequent better selection of plants increased the productivity growth from learning and imitation. While this may be an underlying driver of technology growth, we only find a very weak correlation between contemporaneous changes in aggregate technology and the evolution of average plant productivity (or other moments of the productivity distribution).

Figure A.13: Evolution of cross-sectional correlation of plant productivity and input share



(A) Balanced Sample (B) Survivors (by cohort)

Notes: Input shares are computed based on a Cobb-Douglas aggregator as explained in the text. For within-industry results, we first estimate the correlation across plants in a given industry and year and then construct the weighted average correlation across industries using the industry's average share in manufacturing value added as a time-invariant weight. All series are normalized by the first year.

The increase in technology growth may also be in line with the recent theory in Ottonello and Winberry (2023) whereby constrained firms initially invest in factor accumulation and only later in activities that increase productivity. Given that we find little empirical evidence for capital deepening, only little plant-level labor deepening in the sense of rising labor shares and strong increases in financial access in the run-up to the Asian Financial Crisis, we are rather skeptical that this mechanism can explain large changes in aggregate technology. Still, given the limitations on the capital series and limited evidence on plant-level investments in technology, we cannot rule out that this mechanism is an important driving force of technology growth.

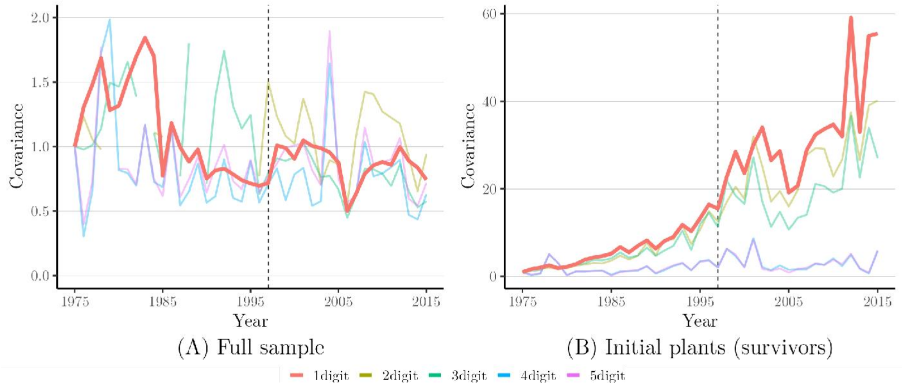
At last, since we cannot distinguish between productivity and demand drivers, changes in demand may also be an important driver of the patterns we observe. For example, the Asian Financial Crisis shows up as a more than 20% drop in technology, which is likely to be at least partly demand-driven. In line with this interpretation, Figure A.12 shows that plant-level exports almost completely plummet in 1998. Decreases in technology may also be partly explained by decreases in demand as the economy grows richer and consumers switch their demand towards services (e.g. Alder et al 2019, Comin et al 2021).

2.A.7 Additional results on changes in misallocation

In this section, we report two sets of additional results. First, we show additional evidence on the evolution of the covariance and correlation of plant-level productivity and input shares. Figure A.13 shows additional evidence on the correlation for a balanced panel of plants and for each cohort of plants between 1970 and 1999. We construct the balanced panel by selecting all plant-year observations between 1975 and 1990 for which the plant operated in 1975 and in 1990 and for which we observe more than 10 observations (to avoid dropping all plants for which individual years are missing or had to be dropped). Extending the time frame would drop too many plants.

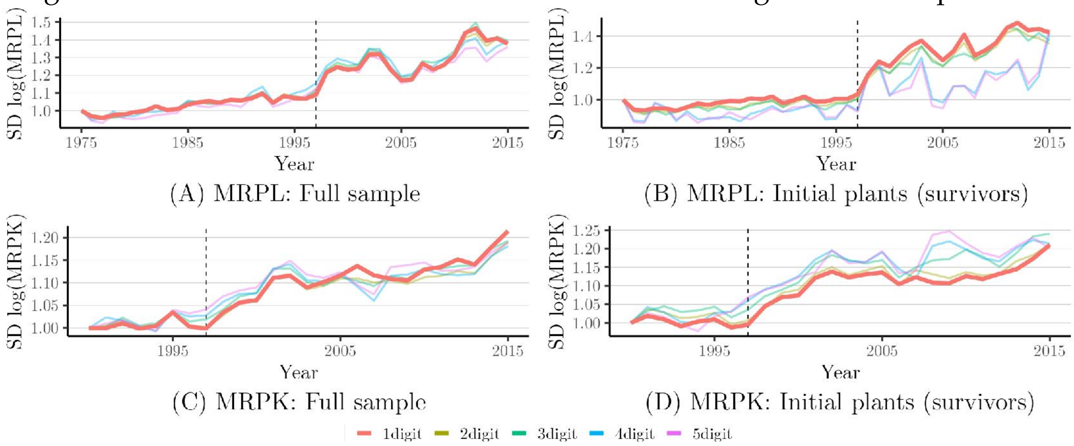
For completeness, Figure A.14 also plots the covariance instead of the correlation. For the full sample, this covariance also does not increase over time due to the entry of small plants. For surviving plants, the covariance increases strongly. This is mechanical, because the sample

Figure A.14: Evolution of cross-sectional covariance of plant productivity and input share



Notes: Input shares are computed based on a Cobb-Douglas aggregator as explained in the text. For within-industry results, we first estimate the covariance across plants in a given industry and year and then construct the weighted average covariance across industries using the industry's average share in manufacturing value added as a time-invariant weight. All series are normalized by the first year.

Figure A.15: Evolution of cross-sectional variation in marginal revenue products

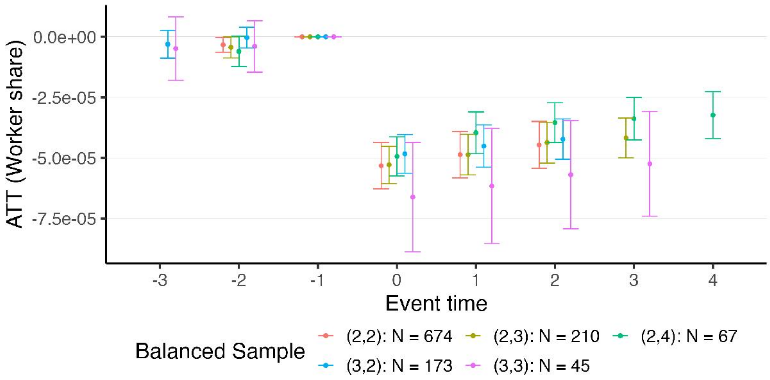


Notes: Evolution of cross-sectional standard deviation in marginal revenue products of labor and capital following Hsieh and Klenow (2009) and Gopinath et al (2017).

shrinks over time, which naturally leads input shares to increase. Also, average productivity strongly increases among surviving plants, adding an additional trend. The correlation is robust to such common trends.

The second set of results is on an alternative measure of the allocation of resources. Figure A.15 reports changes in the dispersion of marginal revenue products of labor and capital; a sign of misallocation in the static model of Hsieh and Klenow (2009). Most importantly, we do not find any evidence for a decreasing dispersion in marginal revenue products over time, which could be linked to an "undoing of misallocation" that drives economic growth. Instead, we find evidence for an increase in the dispersion over time. Most of these increases happen after the Asian Financial Crisis in 1997. We think that the measured increases in the dispersion of marginal revenue products are at least in part driven by changes in measurement. We refer the interested reader to the discussion of measurement changes further above.

Figure A.16: Further event study results for worker shares



Notes: Worker share measured as ratio of number of workers over value added. Treatment definition as in main event study.

2.A.8 Further event study evidence

In this section, we report event study results for hiring responses (in contrast to the labor share responses). We stick to the same treatment definition as previously. To study the dynamic hiring responses of plants to a permanent positive productivity shock, we look at the *worker share* (# worker / value added) instead of the *labor share* (wage bill / value added). The concern of looking at the labor share is that a positive productivity or demand shock at the plant level may lead workers to bargain for higher wages to share in the profit gains. If these bargaining gains slowly accumulate, then we misattribute slow increases in the labor share to labor adjustment frictions. Figure A.16 shows that this concern is unwarranted. Plants actually slowly increase hiring, analogously to the labor share.

2.B Model and Estimation

2.B.1 Adjustment costs as costs of managerial time

In the following, we show that adjustment costs can be microfounded as costs of scarce managerial time. Our goal is to make explicit how adjustment costs can capture the time constraints of a manager working at a plant and to show why it makes sense to write adjustment costs in terms of the costs of labor w_t.

Suppose a plant owner solves the following problem:

$$V(s_{i,t}, h_{i,t-1}, \Omega_t) = \max_{h_{i,t}} \Big\{ y_{it}(s_{it}, h_{i,t}; z_t) - w_t h_t - w_t T(h_{i,t}, h_{i,t-1}) \\ + \lambda(s_{i,t}, h_{i,t}, \Omega_t) \left\{ -\mathbb{E}_c[c_F|\text{stay}] + \beta \mathbb{E}\big[V(s_{i,t+1}, h_{i,t}; \Omega_{t+1}) | s_{i,t}, h_{i,t}; \Omega_t\big] \right\} \Big\}$$

where $T(h_{i,t}, h_{i,t-1})$ gives the managerial time needed to implement changes in the workforce. We denote $T(h_{i,t}, h_{i,t-1})$ in terms of the efficiency units of a worker such that we can express the manager's time cost in terms of the wage w_t. One can think of $w_t T(h_{i,t}, h_{i,t-1})$ as the actual compensation that managers receive or as a combination of compensation and opportunity costs of managerial time.

Time costs are due to two main managerial tasks: (1) the task of hiring and firing, and (2) the task of reorganizing production. Conditional on the task, we think of a single unit of the task as requiring always the same amount of time (e.g. signing one contract always takes a fixed amount of time), but the total units needed depends on the organization and the amount of hiring. Hiring and firing requires (c_F^+, c_F^-) units of time for each unit of labor hired or fired $\Delta h \equiv |h_{it} - h_{it-1}|$. This time comes from filling out paperwork, signing the contracts and adding the worker to the books. Policies that affect the paperwork that plants need to fill out, will change these costs.

Next, for any workers the plant hires or fires, managers need to assign and explain changes in worker tasks. Both for hiring and firing, we assume that managerial time to assign and explain new worker tasks is proportional to the percentage change in the workforce. We assume this is for different reasons in the case of hiring and firing and thus the unit cost of changes in the workforce for hiring and firing can differ. For hiring, the plant hires $h_{it} - h_{it-1}$ workers who they need to explain their new task. The proportional time cost in the case of hiring comes from the possibility that each new worker can also learn from their coworkers. However, in the case of relatively many new hires, each new hire can learn from relatively fewer coworkers, thus increasing the time that the manager needs to add. In the case of firing, jobs may potentially not simply disappear, but need to be reorganized. In this case, the managerial time costs of reorganizing scale with the number of lost jobs $h_{it-1} - h_{it}$ (for which replacements need to be found) and are proportional to the percentage change in the workforce because this is the amount of time the manager needs to reexplain jobs to all existing workers.

In the end, the entrepreneurial time costs T_h are then given by the functional form reported in the main text:

$$T_t(h_{i,t-1}, h_{i,t}) = \begin{cases} c_{0,t}^+(h_{i,t} - h_{i,t-1}) + \frac{c_{1,t}^+}{2}\left(\frac{h_{i,t}-h_{i,t-1}}{h_{i,t-1}}\right)^2 h_{i,t-1} & \text{if } h_t > h_{t-1} \\ 0 & \text{if } h_t = h_{t-1} \\ c_{0,t}^-(h_{i,t-1} - h_{i,t}) + \frac{c_{1,t}^-}{2}\left(\frac{h_{i,t}-h_{i,t-1}}{h_{i,t-1}}\right)^2 h_{i,t-1} & \text{if } h_t < h_{t-1} \end{cases} \tag{2.20}$$

Note that the main text also features fixed costs that can easily be rationalized by fixed time costs of managers whenever there is a change in the organization.

2.B.2 Microfoundation of working capital constraint

The microfoundation of the working capital constraint can be derived from a standard limited enforcement problem (e.g. as in Buera and Shin 2013). Assume that plant managers need to first pay their workers before being able to produce and they do so by borrowing the entire wage bill $w_t h_{it}$ with a financial intermediary. For simplicity, suppose further that the time between production and paying the wage bill is $\varepsilon \to 0$ such that the costs of borrowing go to zero. Suppose further that the plant manager – after paying their workers and producing – could decide to run away with a fraction $\frac{1}{\bar{\kappa}_t}$ of the borrowed resources $w_t h_{it}$. Isomorphically, the plant manager runs away with all of the resources, but is caught with probability $\frac{1}{\bar{\kappa}_t}$. We assume that the only punishment in case of successful evasion is that the financial intermediary can now sue the plant manager and claim (part of) the output of the plant in period t. We assume that the claim is proportional to plant output net of value-added tax. Importantly, the plant manager never loses access to the plant and is not excluded from any future economic activity, ensuring that the constraint remains a static problem. In equilibrium, the financial intermediary will lend $w_t h_{it}$ only to the extent that no plant manager will renege on the contract, implying the financing constraint:

$$w_t h_{it} \geq \kappa_t y_{it}$$

2.B.3 Stationarized value function and balanced growth path after 2015

After 2015, we assume that plants expect wages, all costs and aggregate productivity to rise at the same growth rate $(1+g)$ over time. This allows to capture realistic future growth in a parsimonious way and is in line with the entire economy being on a balanced growth path after 2015. An alternative would be to only enforce constant growth in costs and productivity and then solve for the actual endogenous wage path after 2015 that clears labor markets after 2015. This would require further assumptions on how other fundamentals in the economy evolve (e.g. wedges and technology in the rest of the economy and aggregate labor supply) and feature a continued transition towards an eventual balanced growth path (as long as assumptions on the changes in future fundamentals allow for a balanced growth path). Given that the growth path after 2015 is not identified, we think that our approach strikes a good balance between realism and parsimony.

The value function in 2015 (denoted by T and suppressing dependence on Ω_T for expositional clarity) writes:

$$V_T^M(s_{i,T}, h_{i,T-1}) = \max_{h_{i,T} \in [\underline{h}, \bar{h}]} \Bigg\{ z_T s_{i,T} h_{i,T}^{\theta} k_{i,T}^{\alpha} - w_T h_{i,T} - (r+\delta) k_{i,T} - w_T AC(h_{i,T}, h_{i,T-1}) + \\ \lambda(s_{i,T}, h_{i,T}) \Big\{ -\mathbb{E}_c[c_F | \text{stay}_{i,T}] + \beta \mathbb{E}\big[V^M(s_{i,T+1}, h_{i,T}) | s_{i,T}, h_{i,T}\big] \Big\} \Bigg\} \tag{2.21}$$

In the following, we show that under the right normalization and assuming constant growth for all costs and productivity, we can write: $V_T^M(s_{i,T}, h_{i,T-1}) = \widetilde{V}^M(s_{i,T}, h_{i,T-1}) \tilde{z}_T$, which implies that we can solve for $V_T^M(s_{i,T}, h_{i,T-1})$ by first solving for the stationary $\widetilde{V}^M(s_{i,T}, h_{i,T-1})$ and then renormalizing by $\tilde{z}_T$. To show this, we proceed in two steps. First, we find the normalizing factor $\tilde{z}_T$, which needs to grow at a constant rate $(1+g)$. We do so by deriving the optimal static capital choice: $k_{i,T}^* = \left(\frac{\alpha}{r+\delta} z_T s_{i,T} h_{i,T}^{\theta}\right)^{\frac{1}{1-\alpha}}$. Plugging the capital choice into the value function gives:

$$V_T^M(s_{i,T}, h_{i,T-1}) = \max_{h_{i,T} \in [\underline{h}, \bar{h}]} \Bigg\{ z_T^{\frac{1}{1-\alpha}} s_{i,T} h_{i,T}^{\theta} \left(\frac{\alpha}{r+\delta} s_{i,T} h_{i,T}^{\theta}\right)^{\frac{\alpha}{1-\alpha}} - w_T h_{i,T} - \\ z_T^{\frac{1}{1-\alpha}} (r+\delta) \left(\frac{\alpha}{r+\delta} s_{i,T} h_{i,T}^{\theta}\right)^{\frac{1}{1-\alpha}} - w_T AC(h_{i,T}, h_{i,T-1}) + \\ \lambda(s_{i,T}, h_{i,T}; \Omega_T) \Big\{ -\mathbb{E}_c[c_F | \text{stay}_{i,T}] + \beta \mathbb{E}\big[V^M(s_{i,T+1}, h_{i,T}) | s_{i,T}, h_{i,T}\big] \Big\} \Bigg\} \tag{2.22}$$

from which we can see that for output and capital to grow at a constant rate, we need that

$z_T^{\frac{1}{1-\alpha}} \equiv \tilde{z}_T$ grows at a constant rate. Dividing through by $\tilde{z}_T$ gives the deflated value function $\widetilde{V}^M(s_{i,T}, h_{i,T-1})$:

$$\begin{aligned}\widetilde{V}^M(s_{i,T}, h_{i,T-1}) = \max_{h_{i,T}\in[\underline{h},\bar{h}]} \Bigg\{ & s_{i,T}h_{i,T}^{\theta}\left(\frac{\alpha}{r+\delta}s_{i,T}h_{i,T}^{\theta}\right)^{\frac{\alpha}{1-\alpha}} - \frac{w_T}{\tilde{z}_T}h_{i,T} - \\ & (r+\delta)\left(\frac{\alpha}{r+\delta}s_{i,T}h_{i,T}^{\theta}\right)^{\frac{1}{1-\alpha}} - \frac{w_T}{\tilde{z}_T}AC(h_{i,T}, h_{i,T-1}) + \\ & \lambda(s_{i,T}, h_{i,T};\Omega_T)\left\{-\frac{\mathbb{E}_c[c_F|\text{stay}_{i,T}]}{\tilde{z}_T} + \beta\mathbb{E}\big[(1+g)\widetilde{V}^M(s_{i,T+1}, h_{i,T})|s_{i,T}, h_{i,T}\big]\right\}\Bigg\}\end{aligned} \tag{2.23}$$

where we have made use of the constant growth in $\tilde{z}$:

$$\frac{V_{T+1}^M(s_{i,T+1}, h_{i,T})}{\tilde{z}_T} = \frac{V_{T+1}^M(s_{i,T+1}, h_{i,T})}{\tilde{z}_{T+1}}\frac{\tilde{z}_{T+1}}{\tilde{z}_T} = (1+g)\widetilde{V}^M(s_{i,T+1}, h_{i,T})$$

In the second step, we need to prove that all other aggregate time-varying components also grow at the same rate. In the deflated value function, output and capital do not depend on time-varying aggregates anymore. All terms featuring wages require that wages grow at the same rate $(1+g)$, such that $\frac{w_T}{\tilde{z}_T} = \tilde{w}$ is a constant. The trickier parts are $\lambda(s_{i,T}, h_{i,T};\Omega_T)$ and $\mathbb{E}_c[c_F|\text{stay}_{i,T}]$. We prove that $\lambda(s_{i,T}, h_{i,T};\Omega_T)$ does not depend on time Ω_T if all costs grow by the same rate and that expected fixed costs $\mathbb{E}_c[c_F|\text{stay}_{i,T}]$ grow at the same rate $(1+g)$.

Using the analytic formula for the survival rate, it is easy to see that the survival rate does not vary with time-varying aggregates as long as μ_{xT} and σ_{xT} grow at the same rate $(1+g)$:

$$\begin{aligned}\lambda(s_{i,T}, h_{i,T};\Omega_T) &= exp\left(-exp\left(-\frac{\beta\mathbb{E}\big[\frac{V^M(s_{i,T+1},h_{i,T},\Omega_{T+1})}{\tilde{z}_T}|s_{i,T}, h_{i,T}, \Omega_{T+1}\big] - \frac{\mu_{xT}}{\tilde{z}_T}}{\frac{\sigma_{xT}}{\tilde{z}_T}}\right)\right) \\ &= exp\left(-exp\left(-\frac{\beta\mathbb{E}\big[(1+g)\widetilde{V}^M(s_{i,T+1}, h_{i,T})|s_{i,T}, h_{i,T}, \Omega_{T+1}\big] - \frac{\mu_{xT}}{\tilde{z}_T}}{\frac{\sigma_{xT}}{\tilde{z}_T}}\right)\right)\end{aligned} \tag{2.24}$$

At last, we use the analytic formula for the expected fixed costs to show that they indeed grow at the same rate $(1+g)$:

$$\begin{aligned}\frac{\mathbb{E}_c[c_F|\text{stay}_{i,T}]}{\tilde{z}_T} =& \beta\mathbb{E}\big[(1+g)\widetilde{V}^M(s_{i,T+1}, h_{i,T})|s_{i,T}, h_{i,T}, \Omega_{T+1}\big]\lambda(s_{i,T}, h_{i,T}) - \\ & \frac{\sigma_{xT}}{\tilde{z}_T}\Gamma\left(0, exp\left(-\frac{\beta\mathbb{E}\big[(1+g)\widetilde{V}^M(s_{i,T+1}, h_{i,T})|s_{i,T}, h_{i,T}, \Omega_{T+1}\big] - \frac{\mu_{xT}}{\tilde{z}_T}}{\frac{\sigma_{xT}}{\tilde{z}_T}}\right)\right)\end{aligned} \tag{2.25}$$

which does not depend on aggregate time-varying components.

2.B.4 Formal derivation of main accounting identity

We can start by giving a formal derivation of the main accounting identity that we use to validate our model.

$$\begin{aligned}
Y_t &\equiv \sum_i y_{it} \\
&= \sum_i z_t s_{it} f(x_{it}) = \sum_i z_t s_{it} f(x_{it}) \frac{\sum_i f(x_{it})}{\sum_i f(x_{it})} \\
&= z_t * \sum_i f(x_{it}) * \sum_i s_{it} \frac{f(x_{it})}{\sum_i f(x_{it})} \\
&= z_t * \sum_i f(x_{it}) * \sum_i (s_{it} - \bar{s}_t + \bar{s}_t) \left(\frac{f(x_{it})}{\sum_i f(x_{it})} - \frac{1}{N_t} + \frac{1}{N_t} \right) \\
&= z_t * \sum_i f(x_{it}) * \left[\bar{s}_t + N_t cov \left(s_{it}, \frac{f(x_{it})}{\sum_i f(x_{it})} \right) \right]
\end{aligned}$$

$$ln(Y_t) = ln(z_t) + ln \left(\sum_i f(x_{it}) \right) + ln \left(\bar{s}_t + N_t cov \left(s_{it}, \frac{f(x_{it})}{\sum_i f(x_{it})} \right) \right)$$

2.B.5 Details on estimation

In this section, we provide further details on the model estimation.

Taxes

In the following, we discuss how we map Indonesian corporate income and value added taxes (VAT) over the period 1975-2015 to our model economy. In both cases, we assume that tax rates are constant over time and uniform across firms/plants. This is a very accurate mapping for the VAT, but less accurate for the corporate income tax rate. Throughout, we abstract from the very important topic of tax evasion and enforcement, but we discuss the empirical evidence on this.

We start with the simpler VAT, which we fix in the model economy to a constant 10%, the rate which was officially introduced in 1985 and has remained unchanged in place until now (see: Gillis 1985; Hill 2000; Basri et al. 2021). Officially, the only exemptions are on exports, which we do not model and thus abstract from and there were higher luxury product rates in place that we also abstract from. The VAT replaced an older sales tax that was in place between 1951-1985, whose rates varied from 5 to 20%, but with most sales subject to a 10% rate (Lent and Ojha 1969, 537). Enforcement of the sales tax before 1985 was almost absent, leading to widespread evasion (e.g. see: Gillis 1985), but the introduction of the VAT greatly improved (self-)enforcement and reduced evasion (Hill 2000; Basri et al. 2021), so that our assumption of a flat 10% rate seems reasonable (at least since 1985).

Changes in the corporate income tax rates are slightly more complicated, with major reforms in 1985 and 2009. We simplify our analysis by assuming a fixed 20% corporate income tax rate across firms and over time. Before 1985, many different tax rates were in place, including top marginal rates at 45%, which were non-enforced (Gillis 1985). With the 1985 reforms, corporate income rates were reduced and homogenized, with the maximum marginal rate capped at 35% (Gillis 1985). Between 1985 and 2009, the corporate income tax rates followed a 3-tiered schedule of different marginal tax rates defined over taxable profits (see: Gillis 1985; Basri et al. 2021). The different cutoffs and marginal rates were varied slightly over time, adjusting in part to inflation. For example, in 1985, a tax rate of 15 percent applied to the first IDR 10 million, 25 percent to the next IDR 40 million, and 35 percent on any taxable profits in excess of IDR 50 million (Gillis 1985). By 2009, as documented in Basri et al. (2021), a rate of 10 percent applied for the first IDR 50 million in taxable income; a rate of 15 percent applied for the next IDR 50 million; and a rate of 30 percent applied on all taxable profits over IDR 100 million. After 2009, the corporate income tax system moved to a flat 25 percent rate, with a more complicated schedule of discounts based on gross income that led to effective tax rates below 25 percent (see Basri et al. (2021) for details). Tax evasion and enforcement for the corporate tax rate posed a larger problem than for the VAT, especially before 1985 but also after (Hill 2000, 51f.; Basri et al. 2021). In the end, our 20% flat tax rate assumption tries to parsimoniously capture average effective corporate income tax rates, while abstracting from important temporal and cross-sectional variation.

Estimation of borrowing constraint

To identify the borrowing constraint κ_t, note that the working capital constraint writes as:

$$\frac{w_t h_{it}}{(1-\tau_t^{VAT})y_{it}} \leq \kappa_t$$

The left-hand side of this constraint is a tax-adjusted labor share, which is directly observable. The constraint gives an inequality so that in the absence of measurement error in $w_t h_{it}$ or $(1-\tau_t^{VAT})y_{it}$:

$$\kappa_t = \max_i \left(\frac{w_t h_{it}}{(1-\tau_t^{VAT})y_{it}} \right) \quad \text{for } i \to \infty$$

as long as the constraint is strictly binding for any plant. It is easy to see that the constraint will always bind for some plants as long as there is a non-zero chance for large productivity losses and plants face some positive costs of adjusting labor. The bigger problem is that any estimator based on the maximum observed adjusted labor share will be strongly influenced by outliers and individual measurement error in the labor share. As we discuss in more detail in the data cleaning Appendix, changes in the survey questions over time is one reason why we expect systematic variation in the maximum reported labor share over time that is independent of actual changes in the borrowing constraint κ_t. We thus give up on identifying variation in κ_t over time. Instead, we opt for a robust estimator of κ, taking the 95th percentile of the observed adjusted labor share. Based on this estimator, we find that $\kappa = 1.7$.

Discretization details

We discretize productivity and labor. Specifically, we choose 30 grid points for idiosyncratic productivity, which we select based on quantiles of the productivity distribution. We choose more quantiles at the right tail of the distribution as these high productivity plants are key for the aggregate economy. Non-parametrically estimating the transition matrix of idiosyncratic productivity is quantitatively important as other oft-used processes such as an AR(1) log-normal process cannot replicate empirically observed productivity dynamics (e.g. see Ruiz-García (2019)). We discretize efficiency units of labor h_{t-1} on a grid of 1000 points that we choose based on equal spaced quantiles, ensuring that the entire labor distribution is well represented. Specifically, we choose the bottom 990 grid points based on quantiles (ensuring that all plants in the data can be mapped to the grid) and then use the last 10 grid points to extend the upper bound for labor to allow plants in the model to grow beyond what we observe in the data.

Fundamentals needed for model counterfactuals

Two key sets of model fundamentals are not needed for solving the baseline model, because they are linked to reduced-form statistics that are treated as fixed along the baseline equilibrium path: the path of potential entrant distributions and the fundamentals of the rest-of-the-economy including aggregate labor supply. For model counterfactuals, however, all fundamentals are needed, so we now discuss their identification.

The potential entrant distributions can be related to objects of the baseline equilibrium path: $PE_t(s_t, h_t; \mathbf{\Omega}) = E_t(s_t, h_t; \mathbf{\Omega})/\mathbb{P}_E(s_t, h_t; \mathbf{\Omega})$, where $E_t(s_t, h_t; \mathbf{\Omega})$ is the identified path of entrant distributions and $\mathbb{P}_E(s_t, h_t; \mathbf{\Omega}) = P\big(V^M(s_{i,t}, h_{i,t}; \Omega_t)\big)$ gives the path of entry probability distributions. The latter is a function of the incumbent's value function, which we directly obtain from the baseline model computation, and the entry cost distribution P. Since the potential entrant distributions and the entry cost distribution are not separately identified, we make the identifying assumption that the entry cost distribution is the same as the fixed cost distribution governing plant survival.[31]

The time path of aggregate labor supply is given by the sum of aggregated labor supply in the two sectors of the economy: $H_t = H_t^R + H_t^M$. Total labor supply in manufacturing H_t^M is identified from aggregating up plant-level labor demand h_{it} over the computed equilibrium path. To obtain H_t^R, we use the total observed number of workers l_t^R in the Rest of the Economy and map this to the total efficiency units of labor in R accounting for differential worker selection across sectors in Indonesia.[32]

For the rest-of-the-economy, we can directly identify θ_R and the sequences of A_t and τ_t^R. For this, take plant first-order conditions to obtain: $\frac{\theta_R}{(1+\tau_t^R)} = \frac{w_t h_t^R}{y_t^R}$. We use observed y_t^R and can construct $w_t h_t^R$ to obtain the left-hand side. We assume that wedges behave such that the average of the right-hand side over time is exactly equal to θ_R. Labor wedges τ_t^R are backed

[31]Given that plant entry and survival likely depend on similar economic forces (e.g. similar outside options for not running a plant), we think this gives a reasonable estimate. We also think this gives a conservative estimate of potential entry because most plants survive, implying that most potential entrants also enter. The assumption is a form of normalizing the distribution of potential entrants and is more general than normalizing the total number of potential entrants as often done in entry models (see Aguirregabiria 2021, Chp. 5).

[32]Specifically, we use the estimates of wage differences and worker selection across rural agriculture and urban non-agriculture from Hicks et al. (2017) for Indonesia. This leads us to estimate that average efficiency units of labor are roughly two times larger in M than in R. Hicks et al. (2017), using worker-level panel data from Indonesia, find that non-agricultural jobs earn about 2.5 times higher income than agricultural jobs, but that around 80% of this earnings gap is explained by selection as captured by individual-specific fixed effects. Through the lens of our model, this implies that manufacturing workers have on average more efficiency units of labor. We enforce the point estimates of Hicks et al. (2017) across all time periods.

out such that the previous equation holds exactly. Given θ_R and h_t^R, we can simply back out the sequence A_t using: $A_t = y_t^R / \left(H_t^R\right)^{\theta_R}$.

2.B.6 Details on Euler estimation

In this subsection, we provide more details on the Euler estimation procedure we use and derive all main results.

Derivations for Gumbel distribution

We start out by showing that the Gumbel distribution for fixed costs allows closed-form expressions for the survival probability and the conditional expectation of fixed costs. For expositional clarity, we suppress dependence on the aggregate state Ω_t, but note that all objects generally depend on the aggregate state.

$$\lambda(s_{i,t}, h_{i,t}) = exp\left(-exp\left(\frac{-(x(s_{it}, h_{it}) - \mu_t^x)}{\sigma_t^x}\right)\right) \tag{2.26}$$

$$\mathbb{E}_c[c_F|\text{stay}] \equiv \tilde{g}(s_{i,t}, h_{i,t}) = x(s_{it}, h_{it})\lambda(s_{i,t}, h_{i,t}) - \sigma_t^x \Gamma\left(0, exp\left(\frac{-(x(s_{it}, h_{it}) - \mu_t^x)}{\sigma_t^x}\right)\right) \tag{2.27}$$

where $x(s_{it}, h_{it}) \equiv \beta\mathbb{E}\left[V(s_{i,t+1}, h_{i,t})|s_{it}, h_{it}\right]$ and $\Gamma()$ gives the incomplete Gamma function. That is, in principle, $\mathbb{E}_c[c_F|\text{stay}]$ depends not only non-linearly on the parameters $\{\mu_t^x, \sigma_t^x\}$, but also depends directly on the unknown expected future value x. However, given that the continuation value x is simply an invertible function of (observable) $\lambda(s_{i,t}, h_{i,t})$, we can rewrite the term to substitute for x:

$$\begin{aligned}
\tilde{g}(s_{i,t}, h_{i,t}) =& \mu_t^x \lambda(s_{i,t}, h_{i,t}) - \sigma_t^x \Big\{ ln\left(-ln\left(\lambda(s_{i,t}, h_{i,t})\right)\right)\lambda(s_{i,t}, h_{i,t}) + \Gamma\left(0, -ln\left(\lambda(s_{i,t}, h_{i,t})\right)\right)\Big\} \\
\equiv& \mu_t^x \tilde{g}_1(s_{i,t}, h_{i,t}) - \sigma_t^x \tilde{g}_2(s_{i,t}, h_{i,t})
\end{aligned}$$

We will use these equations and the invertibility of exit rates for continuation values throughout.

Identification details

In the following, we derive the estimating Euler equation and then discuss identification. To derive the estimating Euler equation, we exploit the invertibility of exit rates as shown above and simplify terms to rewrite the Euler equation only in terms of observables and model parameters:

$$\begin{aligned}
0 =& \frac{\partial y(s_{i,t}, k_{it}, h_{i,t}, z_t)}{\partial h_{i,t}} - w_t - w_t \frac{\partial C_h(h_{i,t}, h_{i,t-1}; w_t)}{\partial h_{i,t}} + \\
& \lambda(s_{i,t}, h_{i,t})\beta\mathbb{E}\left[-w_{t+1}\frac{\partial C_h(h_{i,t+1}, h_{i,t}; w_{t+1})}{\partial h_{i,t}}|s_{i,t}, h_{i,t}\right]\Bigg\{ \\
& 1 - \lambda(s_{i,t}, h_{i,t}) + ln\big(\lambda(s_{i,t}, h_{i,t})\big)\left[\frac{\mu_t^x}{\sigma_t^x}(2\lambda(s_{i,t}, h_{i,t}) - 1) - \Gamma\left(0, -ln\left(\lambda(s_{i,t}, h_{i,t})\right)\right) - \right. \\
& \left. ln\big(-ln\big(\lambda(s_{i,t}, h_{i,t})\big)\big)(2\lambda(s_{i,t}, h_{i,t}) - 1) - \lambda(s_{i,t}, h_{i,t})\frac{\partial\Gamma\left(0, -ln\left(\lambda(s_{i,t}, h_{i,t})\right)\right)}{\partial\lambda(s_{i,t}, h_{i,t})}\right]\Bigg\}
\end{aligned}$$

Given the estimating Euler equation, we can now discuss the identification of the parameters. We discuss each of the three sets of parameters in turn.

Linear and convex adjustment cost parameters: Without giving a full identification proof, one can see that the Euler equation generally identifies marginal adjustment costs $\frac{\partial C_h(h_{i,t},h_{i,t-1};w_t)}{\partial h_{i,t}}$ non-parametrically. Given our functional form assumption on adjustment costs, linear costs c_0 & convex costs c_1 are identified as follows: c_0 adds as a fixed wedge between the marginal product and the marginal costs of labor, but any adjustments today save on adjustments tomorrow. Thus, linear costs are pinned down by the observed labor wedge across plants and the probability of switching between shrinking and growing as determined by the volatility of the productivity process. Asymmetric linear costs are identified from the differential behavior of growing and shrinking plants. The convex costs c_1 instead scale with labor growth and are thus identified from the variation in within-plant labor demand growth across periods, again conditioned by the observed volatility of the productivity process. Low labor demand growth despite a high labor wedge will point to strong convex adjustment costs. Again, asymmetry here is identified from differential growth and shrinking (conditional on the productivity process and the state).

Fixed adjustment costs: The Euler equation does not identify fixed costs F^+ & F^- since they do not enter marginal adjustment costs. However, we note that fixed costs are identified from the (time-varying) distribution of plants that are not adjusting and for whom the Euler equation does not hold. The idea is that the more plants choose to not change their labor inputs (as we condition on previous labor and vary productivity), the higher the implied fixed costs. In the data, driven by the choice of focussing on efficiency units of labor, we do not see any plant that remains strictly inactive. We can thus not rule out that fixed costs are zero and fix them to zero throughout. We also note that in a previous version of the paper, we estimated fixed costs indirectly by solving for the model equilibrium path and also found them to be noisily estimated around zero. In cases where one might be particularly interested in fixed costs of adjustment that induce inaction – such as the sluggish responses to aggregate shocks – we think it is best to either work directly with the number of workers or at least work with the nominal wage bill.

Cost parameters (exit): One can immediately see that the Euler equation only identifies the ratio $\frac{\mu_t^x}{\sigma_t^x}$. The reason is that the Euler equation captures the marginal effect of changes in labor demand on the survival probability, which only depends on the ratio of the level and dispersion of costs. What variation in the data identifies this cost ratio? While the dependence in the Euler Equation looks daunting, the cost ratio is jointly disciplined by the size of the labor wedge, the dispersion in survival probabilities and the size of marginal adjustment costs next period over current labor demand. Given empirically estimated survival probabilities, one can see that a higher cost ratio generally increases the labor wedge.[33] Thus, high observed labor wedges push towards lower cost ratios.

[33] This is as long as marginal adjustment costs tomorrow are negative, since $\lambda(s_{i,t}, h_{i,t}) ln(\lambda(s_{i,t}, h_{i,t}))\,(2\lambda(s_{i,t}, h_{i,t}) - 1)$ is generally negative (since survival probabilities are generally higher than 0.5 and $ln(\lambda(s_{i,t}, h_{i,t})) < 0$).

Estimation details

The estimation proceeds in two stages. In the first stage, we estimate reduced-form survival probabilities $\lambda(s_{i,t}, h_{i,t}, \Omega_t)$ and dynamic labor input choices $h(h_{it-1}, s_{it}, \Omega_t)$ conditional on the state space. In the second stage, we enforce these reduced-form objects to estimate the Euler equation for the structural parameters of interest.

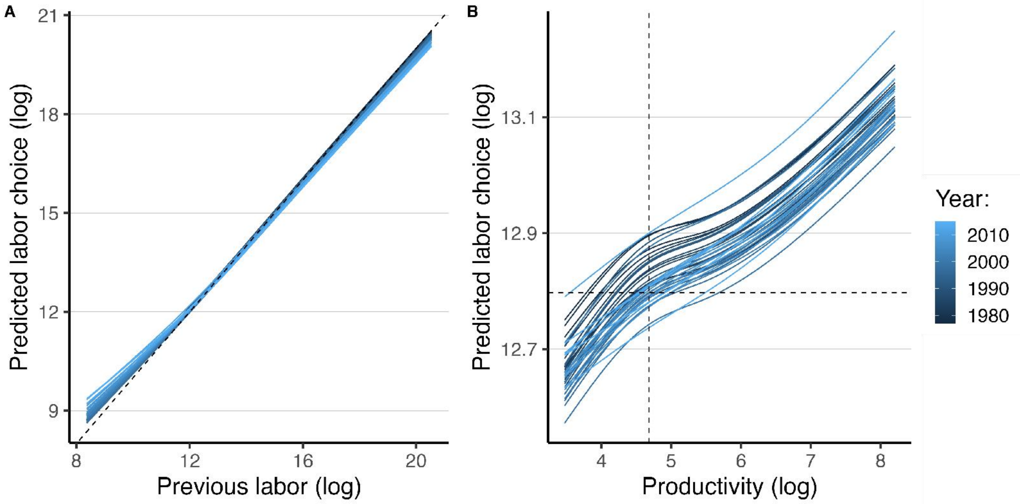
We start with the estimation of conditional survival probabilities and labor input choices. To flexibly estimate both, we draw on generalized additive models for their combination of flexibility and robustness in estimating semi-parametric functional forms. However, subsequent parameter estimates are very similar when using flexible polynomial regressions in the first stage instead. We start with survival probabilities, which – through the lens of the model – are a nonlinear, time-varying function in current labor and productivity. They are time-varying because exit decisions depend on the aggregate state space through current wages and aggregate productivity as well as through perfect foresight over future wages and aggregate productivity. We estimate survival probabilities as the combination of year fixed effects and semi-parametric functions in labor, productivity and their interaction. In general, we find that the estimated survival probabilities make sense while improving on a simple linear model in labor, productivity and year fixed effects. Specifically, the GAM achieves an adjusted R^2 that is roughly 13% higher than the linear model and produces survival probabilities that are increasing in productivity.

For labor input choices, we instead estimate the GAM as the combination of year fixed effects and semi-parametric functions in previous labor, productivity and their interaction.[34] The adjusted R^2 of our GAM is around 96%, driven by the huge explanatory power that previous labor has for current labor. In fact, a simple linear regression of previous labor on current labor already reaches $R^2 = 0.95$ with a coefficient of auto-correlation close to unity. Hence, a policy function that simply says to stick with past labor already explains observed labor choices extremely well. Through the lens of the model and the Euler equation, in the presence of sufficient productivity variation, this already points to large adjustment costs. However, to identify adjustment cost parameters, we require that the estimated policy functions also show variation across productivity, since changes in productivity conditional on previous labor vary the returns to adjusting labor. Figure B.1 plots the policy functions implied by our estimated GAM. Panel A varies previous labor and fixes productivity at the median, while Panel B varies productivity but fixes previous labor at the median. We find that policy functions are monotonic in productivity (in line with the model), but nonlinear such that labor is declining more strongly for low productivities and increasing more strongly for high productivities.

In the second stage, we enforce these reduced-form objects to estimate the Euler equation for the structural parameters of interest. For this stage, we follow the CCP literature in imposing our model structure. This means that we impose the same discretized state space as in our model and the same process of idiosyncratic productivity to ensure that the Euler esti-

[34] We also looked at more flexibility in how policy functions could vary in labor and productivity over time (beyond a simple year fixed effect), however, we found this to give less robust results. The reason is that such higher flexibility is more likely to predict more erratic differences in input choices over consecutive years, which "violate" the smoothing motive implied by the Euler equation and which can then only be rationalized by extreme adjustment cost parameters. We believe that if the underlying data has less measurement issues (that are driven by common aggregate components) and is maybe taken from an economic environment where there are less changes over time, it is more feasible to allow for more flexibility in how policy functions can change over time.

Figure B.1: Step 1 of CCC estimation: flexibly estimated labor input choices



Notes: Panel A gives labor choices over previous labor (fixing productivity at the unconditional sample median). Panel B gives labor choices over productivity (fixing previous labor at the unconditional sample median). Dotted lines give unconditional sample medians.

mation gives parameter estimates that are fully consistent with our model. We also impose CCC-estimated input choices for plants at time t and for all states in $t+1$ conditional on the plant's state in t, greatly reducing the noise in the estimation: this means that for a plant with labor h_{it-1} and s_{it}, we do not use observed h_{it} and h_{it+1}, but instead the predicted values $\hat{h}_{it}(s_{it}, h_{it-1}, \Omega_t)$ and $\hat{h}_{it+1}(s_{it+1}, \hat{h}_{it}, \Omega_{t+1})$, where the latter is over all possible s to be able to compute the expectation term.[35]

We think it is important to briefly mention the computational gains here. An important step in the estimation is that we follow Bajari, Benkard, and Levin (2007) in exploiting the fact that adjustment cost parameters enter linearly in the problem, which is due to the linearity of the marginal adjustment cost specification and due to plants' risk neutrality. Practically, this means that we can compute all expectation terms outside the parameter loop, greatly speeding up the parameter estimation. It is important to highlight the computational gains from this step alone: We can estimate parameters from both steps in less than 2 minutes on a standard personal computer and moving to annually estimated parameters (increasing the number of parameters by a factor of 40) can sometimes even be faster – the reason is that the pre-computed terms all stay the same and instead of estimating X parameters on N data points, we now separately estimate X parameters on N/T data points T times, which one can even parallelize.

We estimate structural parameters via nonlinear least squares (NLS). We do so by assuming that the Euler equation can be written as: $f(\Theta) + \eta_{it} = 0$, where η_{it} is model misspecification error or additive measurement error and Θ is the vector of parameters. Table B.1 reports estimated results. Table B.2 separately estimates parameters for the period before the Asian Financial Crisis and for the period after. One can see that estimated adjustment cost parameters

[35] An alternative to our approach would be to directly use observed plant-level future labor adjustments as noisy realizations of expected labor adjustments, without enforcing model-based expectations (e.g. Hall 1979). Our approach is closer to our model, greatly reducing noise in the estimation, which is particularly problematic for the estimation of convex costs that disproportionately react to outliers. However, this makes our estimation approach – as other CCC/CCP estimators – more susceptible to model misspecification.

Table B.1: Main Euler estimation results

Parameters	Estimates	Std error	95% CI
c_0^+	0.735	0.010	[0.715,0.755]
c_1^+	36.656	0.059	[36.54,36.772]
c_0^-	0.000	0.011	[-0.022,0.022]
c_1^-	12.593	0.073	[12.45,12.736]
Cost ratio	-0.366	0.004	[-0.374,-0.358]

Details: Pooled across all consecutive plant-year observation pairs (N = 358,240). Adjustment cost parameters are restricted to be (weakly) positive and the cost ratio is bounded between -0.577 and -0.366 to ensure that median and mean costs are sufficiently far apart and rationalize dispersion in exit probabilities. Inference for corner solutions should be treated with care. Standard errors are not yet corrected for the multi-stage estimation.

Table B.2: Euler estimation results: Pre vs. Post Crisis

Parameters	Estimates		Std error		95% CI	
	Pre-1997	Post-1999	Pre-1997	Post-1999	Pre-1997	Post-1999
c_0^+	0.635	0.731	0.008	0.013	[0.619,0.651]	[0.706,0.756]
c_1^+	28.590	38.495	0.098	0.069	[28.398,28.782]	[38.356,38.63]
c_0^-	0.000	0.000	0.006	0.013	[-0.012,0.012]	[-0.025,0.025]
c_1^-	7.206	15.605	0.167	0.088	[6.879,7.533]	[15.433,15.777]
Cost ratio	-0.366	-0.366	NaN	0.004	[NaN,NaN]	[-0.374,-0.358]

Details: Pooled across all consecutive plant-year observation pairs (N = 136,918 for pre, N = 209,075 for post). Adjustment cost parameters are restricted to be (weakly) positive and the cost ratio is bounded between -0.577 and -0.366 to ensure that median and mean costs are sufficiently far apart and rationalize dispersion in exit probabilities. Inference for corner solutions should be treated with care. Standard errors are not yet corrected for the multi-stage estimation.

are considerably higher post-1999 than before. At last, we also estimate adjustment costs at an annual level. Figure B.2 shows the yearly estimated convex adjustment costs and shows that they tend to increase over time.

2.B.7 Further model validation exercises

In this section, we show further model validation results. Specifically, Figure C.1 shows the distribution of labor shares. A key feature of the data, which the model captures, is that while average labor shares increase when holding productivity constant, the large observed shifts in productivity due to selection and productivity convergence imply that increasingly more production is concentrated in more productive plants. These productive plants, however, have substantially lower labor shares, in part because they were surprised by positive productivity shocks and adjust labor only slowly, and in part because they avoid large labor increases anticipating future mean reversion in productivity. Together, this implies that the aggregate labor share is low and remains low over time, with the median labor share even declining in the data. Apart from the decline in the median labor share, the model correctly captures these

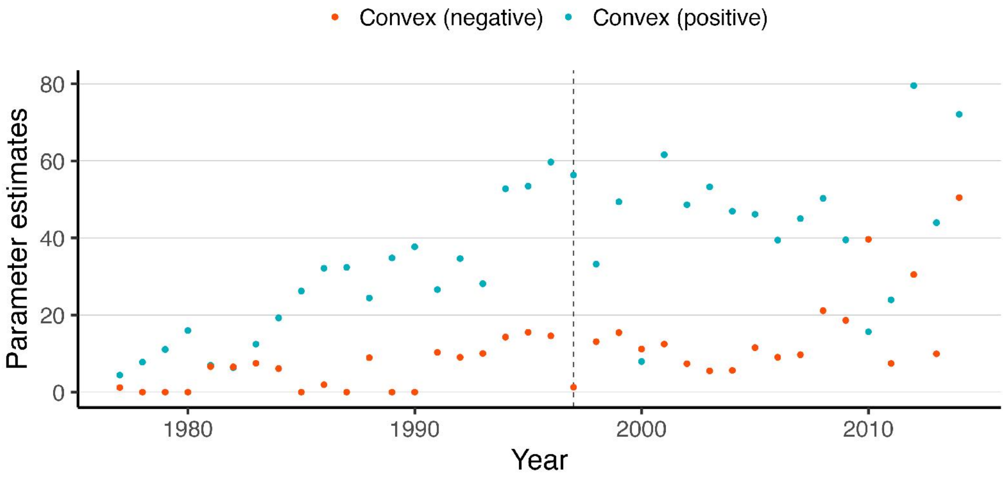


Figure B.2: Annually estimated convex adjustment costs. Details in text.

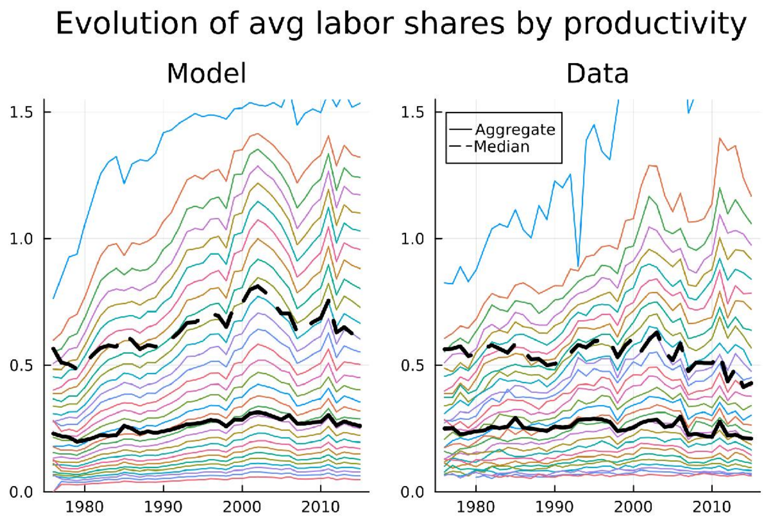


Figure C.1: Changes in the entire distribution: baseline model versus data.

distributional changes.

2.C Counterfactuals and results

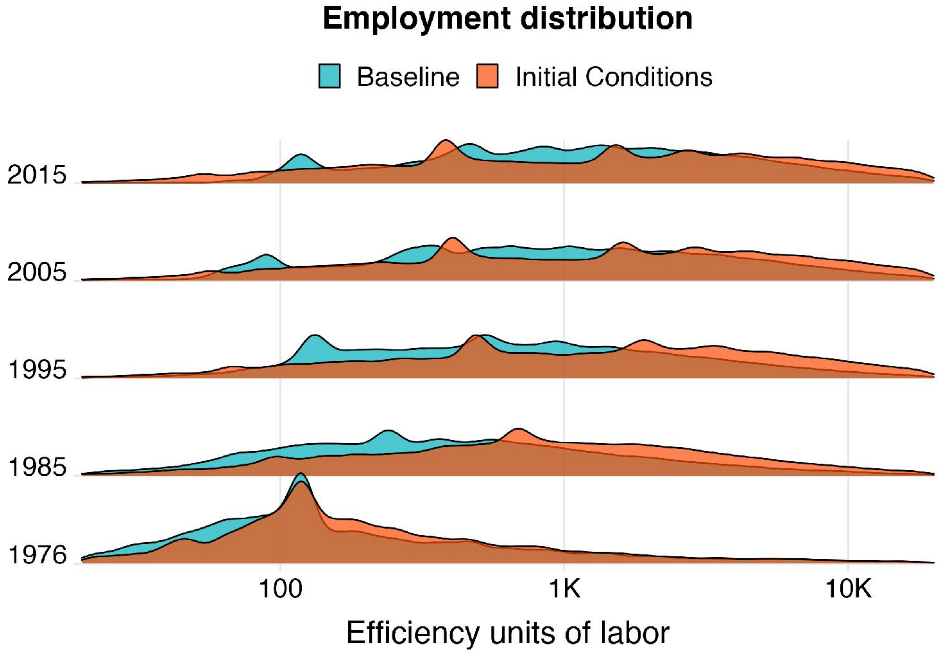
2.C.1 Further results on decomposing the drivers of growth

Figure C.2 reports the evolution of the employment distribution in the baseline (model) economy versus the counterfactual (model) economy where only initial conditions in 1976 play out over time (and all other fundamentals in 1976 are fixed).

2.C.2 Details on INPRES evaluation

This section provides further details on the model-based evaluation of the INPRES school construction program. We start out with more details on how we interpret the program through the lens of the model, how the program's effects map into changes in model fundamentals and

Figure C.2: Evolution of employment distribution: baseline model versus initial conditions counterfactual



Notes: Details in the text.

how we identify counterfactual fundamentals of the economy had the INPRES school construction program not been implemented. In the second part, we then provide more detailed results on the model-based evaluation of the INPRES program.

We assume that the program's direct effect only goes through improving children's education as measured by human capital h in the model. We can be agnostic about how schools raised human capital, capturing a combination of changes along the extensive margin (some children are induced into going to school in the first place), intensive margin (some children stay longer in school) and quality margin (more schools and teachers meant smaller classroom sizes and closer proximity that may provide additional time for learning). We assume that overall demographic changes are not affected. The direct effects of changes in human capital then induce a number of endogenous changes in the model. Specifically, the increase in human capital puts downward pressure on wages and drives up labor demand both in the rest of the economy and across all manufacturing firms. Increases in human capital also have an effect on the endogenous entry and exit of firms, but we assume that this only happens through input costs. Specifically, we rule out that increases in education may have a direct effect on entrepreneurial choices and the distribution of potential entrants. Given that the policy only had measurable effects on primary and secondary schooling outcomes and that entrepreneurs in larger manufacturing firms are more likely to have tertiary education, we think this is a reasonable assumption.

Formally, we model the INPRES program as changing individual-level human capital, which aggregates up to aggregate human capital H_t over time. While the model allows for individuals being differentially affected by the school construction program and also differentially select into different sectors, the general equilibrium results only depend on the change in aggregate human capital. Denoting by L_t the evolution of the number of workers (which we assume to be unaffected by the program), we need to know the average effect of the program on human capital per worker. Through the lens of the model, the average treatment effects estimated via differences-in-differences exactly capture the average differences in human capital h induced by the INPRES program for workers who were treated by the program, netting out aggregate changes in the wage. We can rewrite the effect of the overall program on wages as the combination of three separate effects that have been estimated in the literature: the effect of the program on school construction, school construction on years of education and years of education on wages. We further simplify the setup by assuming that we can treat the three terms as separate expectations (which is true under homogeneous treatment effects or when the shocks driving variation in the treatment effects are independent):

$$\begin{aligned}\mathbb{E}\left[\frac{\partial \text{wage}}{\partial \text{program}}\right] &= \mathbb{E}\left[\frac{\partial \text{wage}}{\partial \text{years of schooling}}\frac{\partial \text{years of schooling}}{\partial \text{no. of schools}}\frac{\partial \text{no. of schools}}{\partial \text{program}}\right] \\ &= \underbrace{\mathbb{E}\left[\frac{\partial \text{wage}}{\partial \text{years of schooling}}\right]}_{\approx 0.1 \text{ (Chaisemartin \& d'Haultfoeille '18)}} \underbrace{\mathbb{E}\left[\frac{\partial \text{years of schooling}}{\partial \text{no. of schools}}\right]}_{\approx 0.25 \text{ (Akresh et al '21)}} \underbrace{\mathbb{E}\left[\frac{\partial \text{no. of schools}}{\partial \text{program}}\right]}_{1.98 \text{ (per 1k children; direct measure)}}\end{aligned}$$

Following Akresh, Halim, and Kleemans (2023), we assume from this that the program on average increased years of schooling by half a year for individuals of any treated cohort. We further follow Akresh, Halim, and Kleemans (2023) by assuming that individuals join the workforce at age 18 and that differences in human capital induced by the program are constant over a person's life, in line with one-time educational gains. The primary schools built by the INPRES program between 1973-1979 are for children between the ages of 7-12 years, such that children fully treated by the INPRES program first joined the labor force by 1984. As in the existing literature, we assume that all cohorts born after 1968 benefit from the INPRES program. This assumes that the last cohort that we observe in 2015 still benefited from INPRES schools in 2009 (their last year of primary school).[36] To avoid having to deal with partial treatment, we further assume that cohorts before 1968 did not benefit from the program. Through the lens of our model, the INPRES program thus led to variation in aggregate human capital over the period 1986 to 2015.

We use the following steps to construct counterfactual paths of aggregate human capital in the absence of the INPRES program:

1. Start from aggregate human capital H_t given by model
2. For each year t between 1975 and 2015:
   - count share of working population affected by INPRES treatment (ϕ_t^T) & get aggregate human capital without INPRES: $\tilde{H}_t = H_t * (1 - \phi_t^T) + H_t * \phi_t^T * \frac{1}{\tilde{\beta}}$ where $\tilde{\beta}$ is the corresponding average treatment effect of the program (here: assume that this is 1.05 given as above)

We thus implicitly assume that all cohorts have the same average human capital. This is unlikely, but in the absence of better worker-level estimates of human capital, this is the best we can do. Still, we think that there are two biases that push in opposite directions so that we

[36] The program initially planned for the INPRES-built schools to last for 20 years, however, Akresh, Halim, and Kleemans (2023) note that most even exist 40 years later. Our assumption implies that the maximum age for an INPRES school in our data is 36 years, well in line with the age range of INPRES schools.

think the overall bias may not be too strong. First, younger cohorts likely have more human capital, which means we overestimate human capital in the absence of the INPRES program $\tilde{H}_t$. At the same time, individuals likely experience human capital increases over their life cycle such that young cohorts have less experience and less human capital, biasing our results in the opposite direction.

To construct the path of ϕ_t^T we draw on representative and harmonized population census data that we retrieve via IPUMS. For each available census wave $t \in \{1980, 1985, 1990, 1995, 2000, 2005, 2010\}$, we construct the share of working age individuals (between 18-65) who have been born in 1968 or later, which gives ϕ_t^T. For the years in between, we extrapolate from the previous census wave assuming there is no differential mortality risk. The treated share in the working population is zero before 1986 and then - due to baby-boom cohorts - increases rapidly to almost 20% by 1990, 50% by 2000 and 75% by 2015.

Applying this time path, we find that the INPRES program raised the annual economy-wide level of human capital by 3.6% by 2015. While important, this effect only accounts for less than 2% of the more than doubling of human capital per worker that we estimated over the entire time period from 1975-2015. These numbers also explain the model-based aggregate effects of the INPRES program that we find.

2.C.3 Details on FDI policy counterfactual

In this section, we provide further details on the model-based evaluation of attracting more foreign-owned plant entrants in Indonesian manufacturing. We start out with more details on the entry of foreign-owned plants, how we interpret this variation and map it to the model, how regulatory changes on FDI map to changes in model fundamentals and how we identify counterfactual fundamentals of the economy had Indonesia's FDI policy been different. In the second part, we then provide more detailed results on the model-based evaluation.

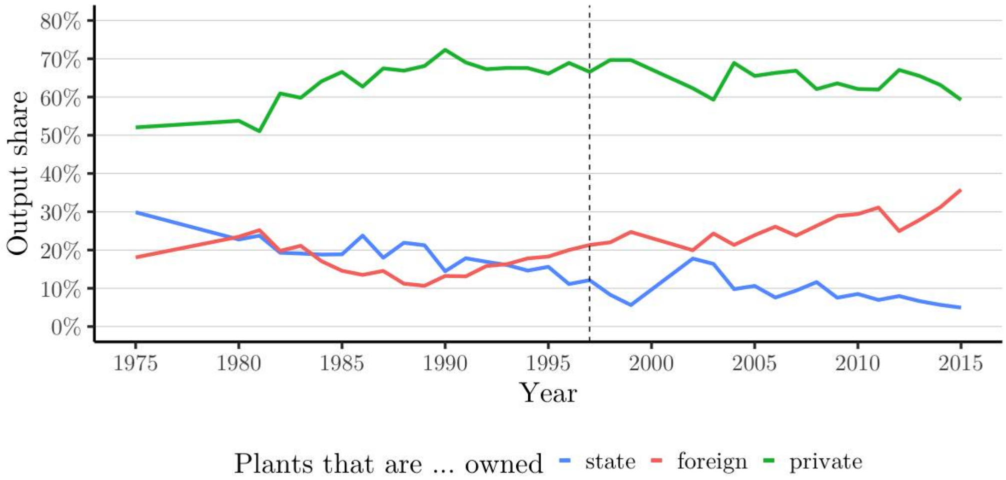
We start out by showing variation in ownership across Indonesian manufacturing plants over time. In Figure C.3 we plot the share of total manufacturing output that is owned either by the state (central + local govt), private domestic or foreign owners. We construct this by summing up all value-added output across plants but taking plants' reported ownership shares as weights. The main movement can be observed in the decline of state ownership from around 30% in 1975 to around 5% in 2015 and the rise of foreign ownership. The domestic private sector is by far the largest actor and owns between 60-70% of all manufacturing production. If we were instead to look at the share of plants, we find that more than 90% are fully domestically owned, which is stable over time. Again, we find that state ownership declines over time and foreign ownership increases, making up almost the entire remainder of 10% by 2015.

For the role of FDI policy, we are specifically interested in the effect on plant entry. Plant entry is particularly important, because most variation in foreign ownership shares is across and not within plants as plant-level ownership shares are relatively constant.[37] Figure C.4 thus reports evidence on the importance of foreign ownership among new entering plants.[38]

[37]For example, the variation in foreign ownership explained by plant fixed effects is 78% and the persistence in foreign ownership as measured by an AR(1) regression is $\rho \approx 0.9$. Restricting only to plants that were ever foreign owned gives slightly lower numbers with the $R^2 \approx 0.6$ and $\rho \approx 0.77$.

[38]We define new entering plants as plants that enter the panel for the first time. While foreign-owned plants are larger and unlikely to not make the cutoff of 20 workers, we want to make sure to compare foreign- and non-

Figure C.3: Evolution of ownership shares for Indonesian manufacturing plants



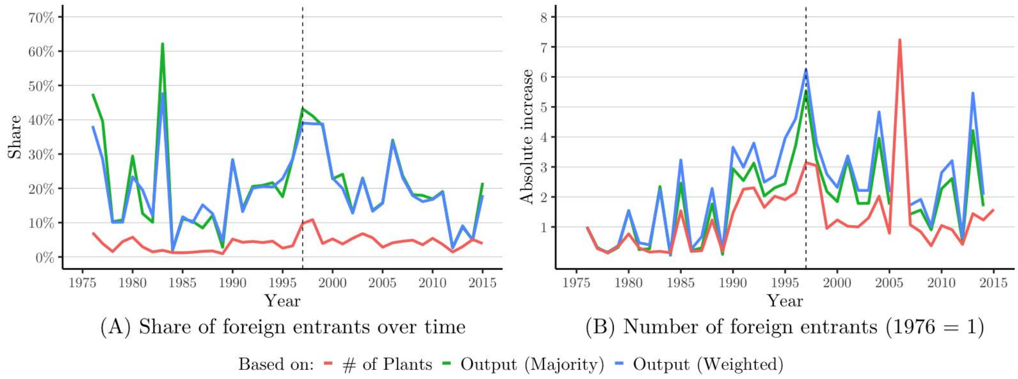
Notes: Data based on Indonesian manufacturing plant census (1975-2015) with 20+ workers. Figure reports the fraction of total manufacturing output that is owned by the state (either local or central government), domestic private owners or foreign owners measured by summing up all value-added output of plants and weighting plants' output by their respective reported ownership shares.

Within a given year, foreign-owned plants (those that are majority foreign owned) make up around 3.8% of entering plants, but they account for 18-19% of total output among entrants. Entrants with some foreign ownership are almost always close to fully foreign-owned with average ownership shares around 80% and the median at 95%. Figure C.4 also documents important variation across time - variation that we exploit for identifying the effect of FDI policy. Specifically, the Indonesian FDI regulatory regime turned increasingly restrictive throughout the 1970s, forbidding 100% foreign ownership and banning FDI entirely in some sectors of the economy (see: Hill 2000). This policy regime reverted only in the second half of the 1980s with simplifications and more transparency over existing restrictions introduced in 1987. In 1992, 100% foreign ownership was permitted again and the 1990s saw increasing attempts at luring foreign manufacturing plants. We can see some of these changes in observed entry, including a marked increase in the absolute and relative weight of foreign entrants between 1990 and the Asian Financial Crisis in 1997. We also see a correlation of these policy changes with the aggregate ownership series, with the importance of foreign ownership increasing steadily from 10% in the late 1980s up to 35% by 2015.

We exploit this policy variation to study the influence of changes in foreign direct investment policy in the Indonesian growth experience. We proceed as follows: We first identify the distribution of actual foreign entrants, which - through the lens of the model - is a reduced form object that masks the underlying distribution of potential foreign entrants. To take out the variation in foreign entry that is purely explained by changes in the economic conditions that make entry more or less attractive, we proceed as before and use the model-identified, time-varying entry probabilities to invert for potential foreign entry distributions. In the next step, we are interested in whether FDI policy changes can account for changes in these potential foreign entry distributions over time. To do so, we compare the period of "restrictive FDI" from 1975-

foreign-owned plants correctly. To this end, we further impose the restriction that the plant has to be younger than 10 years (which is the spacing of the censuses). The sample of entrants without the age restriction looks very similar.

Figure C.4: Evolution of the entry of plants with foreign ownership



(A) Share of foreign entrants over time

(B) Number of foreign entrants (1976 = 1)

Notes: Data based on Indonesian manufacturing plant census (1975-2015) with 20+ workers. In both Panels, the output measures aggregate up value-added across plants, with Majority summing up output for entering plants who are at least 50 percent foreign-owned, and Weighted summing up output based on the respective ownership share. Number of plants instead constructs the share and absolute increase based on the total number of majority-foreign-owned plants.

1986 with the period of "FDI promotion" from 1987-1997. Given the known data limitations of annual variation in entry (e.g. the bunching of entry around census years), we aggregate the potential entry distributions within each of the two time periods. Luckily, the census waves do not fall in a year between, so that there is no ambiguity in how to attribute entrants. We then compare changes in these two aggregated potential entrant distributions. For simplicity, we measure the effect of changes in the FDI policy on changes in potential foreign entrants by comparing the weighted mass of potential entrants across the two periods, taking as weights either the plants' value added or employment at entry.[39]

2.C.4 Making a (more impressive) Growth Miracle

In this part of the Appendix, we move away from Indonesia's historical growth experience and ask whether and how Indonesia could have experienced a more impressive manufacturing growth miracle, closer in comparison to the experiences of countries such as China or Malaysia. For this, we study two important policy levers that both have sizable growth effects, but play out differently over time. Specifically, we look at reduced-form policy changes that either reduce (convex) labor adjustment frictions or increase the annual growth in aggregate technology in manufacturing, but deliver the same long-run growth in manufacturing output.[40]

[39] We thus only use the time series variation and not a differences-in-differences identification design. The model-based trend correction should ensure the validity of the approach and we do not think that it is credible to compare foreign entrants to domestic entrants in a differences-in-differences design, because changes in domestic entrants (e.g. due to changes in demographics) may likely show very different trends. An alternative with a DiD design would be to compare potential entrant distributions across different countries or across different industries that were differentially treated by the regulatory regime. This is an interesting approach that we leave for future work.

[40] To reduce labor adjustment frictions, we consider a policy package that reduces both linear and convex adjustment costs. For the linear hiring and firing costs, we consider a hiring subsidy for each new hire of around 25% of the annual wage bill. Changes in convex costs are harder to map directly to tangible policies. Given our microfoundation in terms of scarce managerial time and talent, we think of them as policies that improve managerial quality in the economy such as training programs. We consider a feasible policy mix that halves the estimated convex cost parameters (for hiring and firing) – in line with the lower end of annually estimated adjustment cost parameters that we find in the data. For aggregate technology in manufacturing, z_t, we consider a policy that raises its annual growth by a constant rate to the degree that manufacturing output in 2015 is the same as in the adjustment cost

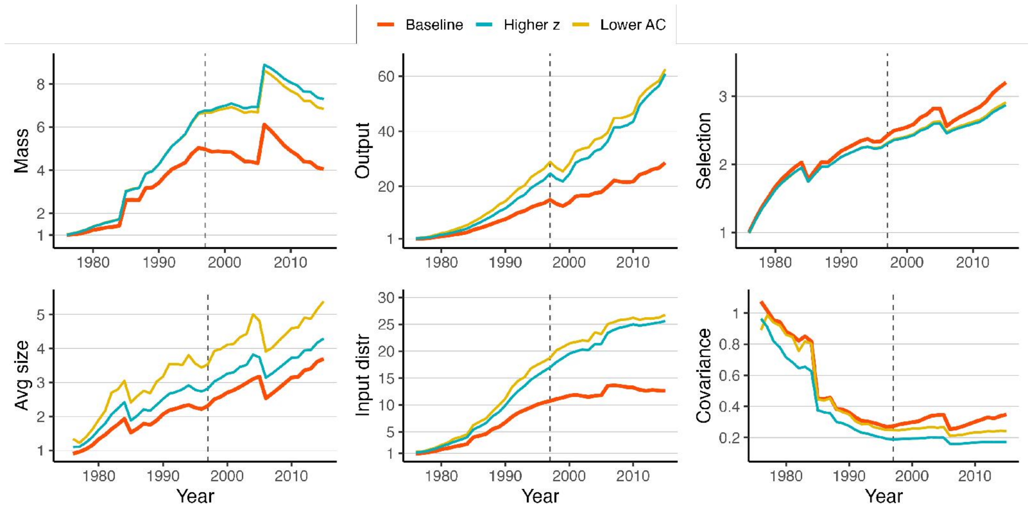


Figure C.5: Main miracle economy counterfactuals.

Figure C.5 shows how the manufacturing miracle would have played out differently in the two alternative scenarios that both see a doubling of manufacturing output by 2015 compared to the baseline miracle economy. As expected, lower adjustment costs lead to faster hiring and thus faster transitions such that output growth is initially higher. With lower adjustment costs, far more large manufacturing plants emerge, driving up the average plant size. Growth in manufacturing technology, on the other hand, makes all plants more productive, leading to more entry and less exit of small plants, a stronger left tail and a much lower covariance of idiosyncratic productivity and resources.

counterfactual.

Chapter 3

Dysfunctional Firm Dynamics and Mexico's Dismal Productivity Performance

Oscar Fentanes[1] & Santiago Levy [2]

Abstract

Over the last three decades, Total Factor Productivity growth in Latin America has disappointed and informality persisted. To shed light on this outcome, we exploit a unique database (by Latin American standards) for Mexico, a country where manufacturing exports grew from seven to 33 per cent of GDP, but labor informality barely changed, firm informality increased, and TFP growth was negative. We construct a twenty-year panel and analyze firm dynamics from two perspectives, the formal-informal and the sector composition of the economy. In the first case we show that high productivity formal firms exited; surviving firms hardly grew, and their productivity fell because more informalized than formalized; and entrants were less productive than survivors, mostly because of large informal entry. In the second case we show that while manufacturing performed relatively better than services and commerce, its contribution to TFP was modest because informality persisted in this sector; and that despite spectacular export growth, the country de-industrialized. We document that for TFP, the formal-informal composition of the economy is more important than its sector composition. While our insights are based on Mexican data, they extend to countries in Latin America and other regions characterized by large informal sectors.

JEL Codes: O17, O11, O54

Keywords: Productivity, informality, misallocation, Latin-America

[1]Toulouse School of Economics; oscar.fentanes@gmail.com.

[2]Brookings Institution; santiagolevy4@gmail.com. We thank Juan Blyde, Mariano Bosch, Matias Busso, Chang-Tai Hsieh, Alain Ize, Jonas Gathen, Marcela Meléndez, Mauricio Mesquita and Dani Rodrik for very useful comments on an earlier draft. We are grateful to Natalia Volkow and the Mexican Institute of Statistics and Geography for facilitating access to confidential microdata. All results have been reviewed to ensure that no confidential information is disclosed.

3.1 Introduction

Barring East Asia, from 1990 to 2019 factor accumulation in Latin America was faster than in any other region of the world, but average annual TFP growth was (-) 0.08 per cent (Fernández-Arias and Fernández-Arias, 2021). In parallel, Latin America was characterized by large and persistent informality (Gasparini and Tornarolli, 2009; Maurizio, 2021). This outcome is puzzling because, with a few exceptions (like Argentina and Venezuela), after the "lost decade" of the 1980s, most countries achieved macroeconomic stability and carried out many reforms to increase efficiency.

Mexico is a salient example of this puzzle. During the 1990s it created various regulatory agencies to increase domestic competition; privatized more than one thousand state-owned enterprises; and signed fourteen free trade agreements with over fifty countries including, notably, one with Canada and the United States. As a result of these reforms, manufacturing exports almost quintupled, from 7% of GDP in 1990 to 33% in 2019. Mexico now exports more manufactures than the rest of Latin America combined. However, between 1990 and 2019, TFP contracted at an annual rate of (-) 0.5%, labor informality barely changed and, as documented below, firm informality increased.

Mexico also provides a good case study to shed light on this puzzle, because it is the only country in the region that for over two-decades has collected data on firms of all sizes and formality status in all sectors. This data allows to follow the patterns of entry, survival, growth and exit of individual firms, measure their productivity and factor shares, and reconstruct the path of aggregate TFP. The study of firm dynamics sheds light on why, despite many efficiency-enhancing reforms, TFP stagnated. In the end, during the period considered, Mexico's economy was subject to two contradictory forces: on one hand, measures to improve efficiency like the ones listed above. On the other, chiefly but not only, flawed tax, labor, social insurance and contract enforcement institutions, that persistently distorted the allocation of capital and labor across firms (Levy, 2018).

In this paper we document that the second set of forces prevailed. After classifying firms by size and formality status, we present stylized facts on resource allocation and market shares in the aggregate and at a very detailed sector level (six-digits of the North American Industrial Classification System, NAICS). We show that firm informality became more widespread, that productivity differences between formal and informal firms increased across manufacturing, services and commerce, and that the distributions of firm productivity and firm size polarized.

Next, we construct a twenty-year panel of firms and extend the Olley-Pakes productivity decomposition proposed by Melitz and Polanec (2015) to study firm dynamics from two complementary perspectives. The first one focusing on the formal-informal segmentation of the economy; the second one on the differences between manufactures, the sector most directly impacted by the trade liberalization measures, and services and commerce. These decompositions lead to the four main conclusions of our paper: first, despite the efficiency-enhancing reforms, informality persisted and was the main proximate reason behind the fall in TFP. Second, while manufacturing experienced productivity gains, these were modest because some informal firms survived, and new ones entered into the sector. Third, despite the fact that services and commerce experienced productivity losses, their share in total resources increased; manufacturing shrank even though it was the higher productivity and better performing sector. And fourth, all-in-all, misallocation increased within and across sectors.

We also take advantage of our panel to focus on surviving firms and study their patterns of growth. We first document that, contrary to expectations, more firms transited from formal to informal status than in the opposite direction. Second, that very few informal firms formalized and became more productive. Third, that while the average size of surviving firms increased, their productivity fell. Finally, we show that calculations of firm growth obtained from firms' age-size profile using data from one period only, as in Hsieh and Klenow (2014), overestimate firm growth. Firms in Mexico hardly grow, particularly medium and large ones.

To the best of our knowledge, this is the first paper that performs a dynamic productivity decomposition for a Latin American country classifying firms by formality status and sector.[3] This allows our paper to relate to four strands of the literature. First, the one associating the firm size distribution and development. Bento and Restuccia (2017) document that there is a positive correlation between aggregate TFP and average firm size; and Poschke (2018) documents that as countries per capita GDP increases, average firms size increases as well and the right tail of the firm productivity distribution thickens. We show, however, that when firm dynamics are dysfunctional, average firm size can increase without aggregate TFP gains, as average size is driven up by the entry of a few large firms, while the survival and entry of small and unproductive firms thickens the left tail of the productivity distribution, driving aggregate TFP down.

Second, we relate to the misallocation literature as in Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). Developing countries, particularly those in Latin-American, present significant misallocation (Pagés-Serra, 2010; Busso et al., 2012; Lederman et al., 2014; Álvarez et al., 2018). While sudden reforms might initiate a process of massive resource reallocation, contradictory forces might act as a bottleneck (Buera and Shin, 2013). Our paper shows that, at least in the case of Mexico, these contradictory forces can not only slow down aggregate TFP but contract it, undoing the benefits of measures to improve efficiency.

Third, we speak to the literature on informality (reviewed by (La Porta and Shleifer, 2014; Ulyssea, 2018)). Static cross-country comparisons suggests that informality becomes less important with development (La Porta and Shleifer, 2014). Our paper, however, shows that informality can persist and, by lowering aggregate TFP, slow down GDP growth. Finally, we speak to the literature on premature 'de-industrialization', as in Rodrik (2016), and document a case where despite a spectacular increase in manufacturing exports, the share of manufactures in GDP falls.

While our paper is based on Mexican data, it offers insights on the relation between sector composition, informality, and TFP, that are likely relevant to other countries in Latin American and elsewhere characterized by institutional arrangements that also result in a large informal sector.

The rest of the paper proceeds as follows. Section 3.2 briefly discusses the institutions generating Mexico's formal-informal divide and their impact on resource allocation. Section 3.3

[3] Eslava et al. (2022) construct a 30-year panel of firms for Colombia. Unfortunately, their data only cover manufacturing and exclude firms with 10 or fewer workers, leaving out most informal firms. Nevertheless, they find that the "up or out" patterns found in the United States are much weaker in Colombia as a result of the survival of small unproductive plants and much weaker selection of new ones; an important result that helps to understand why TFP underperforms in that country, and that is consistent with our findings.

describes the data, the construction of the panel, and the criteria to classify firms. Section 3.4 shows stylized facts on resource allocation and market shares. Section 3.5 discusses our estimates of firm productivity and carries out comparisons across firm size, sector and formality status. Section 3.6 presents the results of the Olley-Pakes decomposition when firms are classified by formality status. Section 3.7 focuses on surviving firms to discuss the relation between firm size, firm growth, and productivity. Section 3.8 presents the Olley-Pakes decomposition when firms are classified by sector. A back-of-the-envelope calculation in section 3.9 shows that in the absence of informality Mexico would have experienced positive TFP growth. Section 3.10 presents our conclusions.

3.2 Brief note on informality and resource allocation

Many institutions in Mexico stand behind the fact that almost 60% of workers and 90% of firms are informal (as defined below), but three stand out (Levy, 2018). First, the legal distinction between salaried and non-salaried workers. The former are hired under a relation of dependency and subordination to work a fixed number of hours in the tasks dictated by a boss/firm, in exchange for a remuneration proportional to the time worked (salary). The latter can work on their own; or be associated with firms but without a relation of subordination, need not work a fixed number of hours, and are remunerated through various schemes: on a piece-meal basis, profit-sharing, or a commission per unit produced or sold.

Firms and workers in salaried contractual relations must jointly contribute to a fixed bundle of social insurance programs including health, pensions, housing, day care, and other benefits. In addition, firms must pay workers at least the minimum wage, cannot dismiss them at will, and when they can, incur in large severance payments.[4] On the other hand, firms and workers in non-salaried contractual relations are not subject to these regulations, and workers can access an unbundled set of health, pensions, day care, and related benefits financed from general tax revenues. The same holds for self-employed workers (one-person firms).

Because workers undervalue the benefits of IMSS, they and the firms that hire them are de facto taxed, generating incentives to evade these contributions and, in parallel, income taxes. On the other hand, non-salaried workers are subsidized because the costs of their social insurance benefits do not have to be internalized in the contract between them and the firm, nor the contingent costs of dismissal regulations. Further, remunerations can be lower than the minimum wage and firms are not obligated to withhold workers' income taxes (Levy, 2018).

Given Mexico's context of imperfect enforcement, some firms hire salaried workers without contributing to IMSS. As a result, the labor force divides into two categories: salaried workers hired legally (formal), and non-salaried workers together with salaried workers hired illegally (informal). Importantly, the latter can access to the same social insurance benefits that non-salaried workers receive, so that the implicit subsidy extends to them and, indirectly, the firms that hire them.

[4] Firms and workers contribute to IMSS (the Spanish acronym for Mexico's social security institute), to Infonavit (the housing institute) and to the Afores (the private administrators of retirement pension funds). Minimum wages are enforced by the Labor Ministry. Dismissals are regulated by labor tribunals. Henceforth we refer to all these agencies as IMSS, in the understanding that this includes Infonavit, the Afores, the Labor Ministry and labor tribunals.

Taxation is the second institution behind the formal-informal divide. Firms pay income taxes under two regimes depending on their annual sales. If they are below approximately US$100,000, firms qualify to a preferential regime where taxes are two percent of sales (under the chapter for individuals). If sales exceed that threshold, firms are in the general regime, where taxes are 30 percent of profits (under the chapter for corporations).[5] Only firms with non-salaried workers can qualify for the preferential regime because those hiring salaried workers must register as a corporation (or cooperative). These asymmetries are accentuated by the fact that firms producing approximately 20% of the consumption basket are exempt from VAT on final sales, and firms producing an additional 26% of that basket are also exempt from VAT on intermediate inputs; and by the fact that firms in the preferential regime cannot issue VAT receipts to firms in the general regime, thus limiting their sales to final consumers or other firms in the preferential regime. The upshot is that firms with non-salaried workers and sales below the threshold, aside from having no social insurance, minimum wage or dismissal obligations, face a very low burden of income taxation and, depending on the good produced, do not have to charge VAT on their sales or pay it on their inputs.

The third institution is associated with the regulation of commercial and credit contracts. Most firms in Mexico, particularly small ones, do not register as a corporation, where the assets of the firm are separated from the assets of the owners; indeed, many are family firms in the sense that owners and workers are relatives, with non-salaried contractual relations between them. On one hand, registering excludes them from the preferential regime of the income tax law, and in any event the costs of doing so are high (transaction costs, notaries). On the other, the benefits, like access to commercial bank credit, may be low because when contract enforcement depends on slow and often corrupt courts, banks substantially undervalue firms' collateral, particularly if they are small, limiting their access to credit.

Considered jointly, these institutions are principally responsible for three outcomes: first, firms' face different labor costs depending on the contractual status of their workers. Second, firms with non-salaried workers or hiring salaried workers illegally are de facto subsidized (as long as they are small), while firms hiring salaried workers legally are taxed. And third, the size distribution of firms is biased towards smallness. As a result, firms with very different productivities can coexist in the same (narrowly defined) market.

In other words, the institutions that give rise to the formal-informal division of economic activity de facto misallocate resources. That said, other institutions also contribute to misallocation in Mexico: the exercise of monopoly power by a few large private firms, uncompetitive and at times corrupt public procurement practices by government agencies, and uncompetitive behavior by state-owned enterprises in the energy sector.

Jointly, during the last three decades the institutions that generate misallocation in Mexico operated in the opposite direction vis-à-vis the efficiency enhancing reforms promoted over the same period. In this paper we do not identify the individual impact of any of these factors on resource allocation and TFP, positive or negative. Rather, we focus on the effect of all of them at the same time, as reflected in the data captured by the Economic Census. Our paper therefore does not focus on causality, but on measuring the net impact of a large number of contradictory policies affecting the behavior of firms and workers during the period studied

[5] Levy (2018) shows that this dual tax regime generates a large discontinuity in firms' after-tax profit functions, implying that increasing sales is not profitable, unless the increase is very large.

here.

3.3 Definitions and data

3.3.1 Definitions

We define formal firms as those that pay at least one peso in social insurance contributions to IMSS. This definition encompasses firms cheating along the extensive margin (not enrolling all of their workers with IMSS), the intensive margin (under declaring their wages), or both. It also encompasses firms mixing salaried and non-salaried workers, as long as they pay something to IMSS for their salaried workers. Further, some firms in Mexico sub-contract some or all of their salaried workers. Unfortunately, the census data does not allow to verify whether firms providing workers to sub-contracting firms in turn comply with their obligations to IMSS. Here we assume that they do, at least partly, and classify firms that sub-contract as formal. Clearly, our definition of firm formality is very generous. But it is appropriate for our purposes because it implies that the firm is registered with IMSS, is subject to labor regulations, is obligated to pay income taxes under the general regime and, when appropriate, can issue VAT receipts on its sales.[6]

There are two types of informal firms. Non-compliant ones, hiring salaried workers but not paying anything to IMSS. And legal ones, those engaged only with non-salaried workers, and thus not required to pay anything to IMSS or comply with regulations on dismissal or minimum wages.[7]

We also classify firms by size, measured by number of workers: very small, 1 to 5; small, 6 to 10; medium, 11 to 50; and large, 51 or more. The classification is attuned to Mexico's context and differs from the one used in other OECD countries, where large firms have at least 100 workers.

Firms' formality status matters for two reasons: social protection and productivity. It matters for social protection because it speaks to the social benefits that their workers are entitled to. But it matters for productivity because it determines firms' flow and contingent costs of labor, their access to institutions in charge of contract enforcement, and sometimes their tax regime. Differently put, formality status impacts critical dimensions of firm behavior like which technologies to adopt, the number of workers to hire and their contractual modalities (including whether to comply fully or partly with the Law), the sources of finance, the range of clients, the ability to adjust to output or technology shocks, and so on.

The formal-informal labels are usually motivated by social protection considerations and can cause confusion when applied to firms. Because our focus here is on productivity, we could avoid them altogether and instead refer to two types of firms. First, those that hire salaried

[6]Busso et al. (2012) separate firms between those that comply fully and partially with their obligations towards IMSS, and those that mix salaried and non-salaried workers. They show that the productivity of these intermediate cases is similar to that of the formal firms defined here, allowing us to use a simpler classification.

[7]Thus, informality is not equivalent to illegality. In fact, as shown below, the majority of informal firms are legal. This differs from other countries, like Brazil, where firm informality implies firm illegality (Ulyssea, 2018). The fact that a large segment of informal economic activity is legal indicates that informality in Mexico is not mostly the result of imperfect enforcement; it is more complex than that and is associated with the institutions discussed in section 3.2.

workers, pay IMSS, are subject to regulations on minimum wages and dismissal and the provisions of the corporate tax regime, and can issue VAT receipts to other firms. And second, those that hire salaried workers but break the Law and pay nothing to IMSS nor observe labor regulations; or have non-salaried workers, do not have to pay anything to IMSS, may pay taxes under the preferential regime but may not issue receipts for VAT, and are not bound by regulations on dismissal or minimum wages. But because this language is more cumbersome, we use the better-known formal-informal labels in the understanding that they are short-hand expressions for the very different circumstances faced by firms.

3.3.2 Data

Every five years, Mexico's statistical institute produces an Economic Census collecting data from firms of all sizes in urban areas operating in a fixed premise (walls, ceiling). Here we use the censuses from 1998 to 2018.[8] The Census classifies firms into sectors at the six-digit level of the NAICS. In the 2018 Census there were 981 sectors, a very detailed level of aggregation which allows to compare the productivity of firms producing very similar goods.

The Census captures a large number of firms: 2.8 million in 1998, 3.0 in 2003, 3.7 in 2008, 4.2 in 2013 and 4.7 in 2018. Importantly, despite its broad coverage, the Census leaves out a substantial amount of economic activity. For instance, the 2018 Census only captures 52% of total employment, an indication of the large number of workers and firms carrying out their activities in mobile premises in the streets of Mexico's cities.[9] In this paper we focus on firms in manufacturing, services, and commerce, which in the 2018 Census represent 98% of all firms, 91% of employment, 66% of capital, and 629 out of the 981 six-digit sectors of the NAICS.

The Census reports the value of the capital owned by firms and the payments made for renting capital goods from other firms. To produce a homogeneous measure of capital, we capitalize payments for rented capital (at 10%) and add them to firms' own capital. In turn, value added is corrected to incorporate payments made by firms for renting capital goods. The Census divides capital into three components: buildings and constructions, transport equipment, and machines. We have price indices for each over the 20-year period considered here and express firms' capital stock in constant prices of 2013.

We measure labor input as the value of payments to people working in the firm, including firm-owners and those hired by honorarium. This measure captures differences in remunerations associated with differences in individuals' schooling and skills. We use the consumer price index to express labor input in prices of 2013. The Census reports the number and payments to workers, but not payments to firm-owners and personnel hired under honorarium. We impute the latter using the median wage of workers in firms in the same six-digit sector.[10]

[8] Unfortunately, the 1993 Census cannot be used because its sector classification differs from the NAICS, which was adopted by Mexico as of the 1998 Census. In the text we refer to firms, although the Census collects data on establishments; that said, 99.7% of firms in Mexico only have one establishment (Levy, 2018).

[9] The Census captures firms in localities of 2,500 or more inhabitants, where over 80% of Mexico's population lives. It excludes firms in smaller localities and in rural areas, and firms in larger localities that do not have a fixed premise: street markets and the like. Employment in rural areas and the public sector represent less than 20% of the total, so urban employment in mobile premises is large, approximately 28% of the total. The point is that there are many more firms in Mexico than captured in the Census, although it is not possible to determine the exact number.

[10] Including firm-owners in labor input is quite important since many informal firms are family enterprises with two to three people including the owner. Workers hired under honorarium are few, but we consider them to better approximate labor input in a context where the contractual structure of firms is heavily influenced by

Finally, we also have price indices for value added at the three-digit level classification of the NAICS. For our TFP decompositions we assign firms into 67 three-digit sectors, 21 in manufacturing, 30 in services and 16 in commerce. Using the corresponding price indices, we compute firm value added in constant prices of 2013.

3.3.3 Panel of firms

Firms in the 2008, 2013 and 2018 censuses have a unique identifier generated by Mexico's statistical institute, allowing to construct a panel for this ten-year period. To extend it back to 1998, we take advantage of the fact that all censuses register firm age, name, legal status, six-digit sector, and detailed location (up to street block).

In a previous paper, we developed an algorithm to match firms in the 1998, 2003 and 2008 censuses based on these characteristics; see Busso et al. (2018). In the simplest case, if a firm in the 2003 census has the same location, legal status, name and six-digit sector than a firm in the 1998 census, and is 5 years older, we consider it to be the same firm.[11]

We evaluate the accuracy of our procedure comparing the results of the algorithm matches between the 2008 and 2013 census with the actual matches using the unique firm identifier given by the statistical institute. Our procedure matches exactly 96% of all firms. Missing matches refer to very small firms, as small, medium, and large ones are matched with 100% accuracy.

In sum, we construct a 20-year panel combining the exact 2008-2018 panel with the 1998-2008 almost-exact panel. We next identify firm exit, entry, and survival over the 20-year period and within each 5-year period. Because the volume of information is extremely large, in what follows we only present the results for the 20-year period and descriptive statistics for 1998 and 2018.

3.4 Stylized facts: resource allocation and market shares

Table 3.1 shows the size and formal-informal composition of firms in 1998 and 2018. Two well-known facts are confirmed. First, the size distribution is strongly skewed towards smallness as 90% of firms have at most 5 workers and less than 1% have more than 50. Second, most firms are informal, and informality is inversely correlated with size.

Two less well-known facts are also shown in Table 3.1: first, more than 60% of all firms are informal but legal; the majority of them very small and not registered as a corporation.[12] Second, firm informality increased between 1998 and 2018, mostly as a result of an increase in

the institutions discussed in section 3.2. We prefer the median rather than the mean wage since the latter can be influenced by outliers. Imputations are done at the six-digit level to reflect as much as possible the specifics of each sector.

[11] The procedure works in most cases, but not all because sometimes there are minor variations in the name. For instance, a firm may appear in the 1998 Census under the name "Muebles de Madera Don Pedro" and in the 2003 one as "Muebles de Madera D. Pedro". In this case, even if the name does not match exactly, we consider it to be the same firm, as long as the other characteristics (age, location, six-digit sector) match. We thank Mexico's statistical institute for giving us access to the detailed firm records. (Muebles de madera stands for wood furniture.)

[12] 88% of firms in the 2018 Census were not registered as a corporation. While the census has no direct information, it is very likely that most of them are family firms, in the sense that owners and workers are relatives, a situation consistent with the absence of salaried contractual relations.

Table 3.1: Firm size and formal-informal composition, 1998 vs. 2018 (Percentage shares)

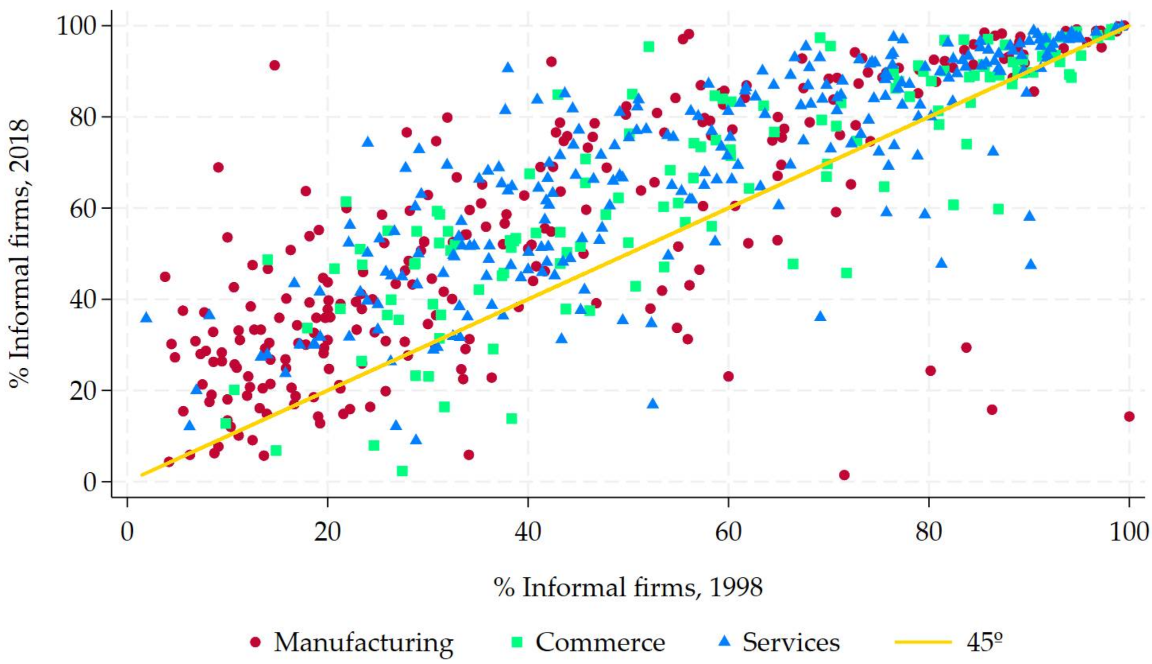
	Formal	Informal and legal	Informal and non-compliant	Total
1998				
1-5	9.84	66.95	13.87	90.67
6-10	3.22	0.56	0.84	4.62
11 - 50	3.22	0.18	0.31	3.72
51+	0.95	0.01	0.03	0.99
Total	**17.23**	**67.7**	**15.07**	**100**
2018				
1-5	5.1	62.93	21.74	89.77
6-10	2.84	0.63	2.16	5.64
11 - 50	2.65	0.25	0.77	3.67
51+	0.87	0.02	0.04	0.93
Total	**11.45**	**63.83**	**24.71**	**100**

Notes: authors' calculations with Census data.

the number of informal and non-compliant firms. In 2018, one out of every four firms hired salaried workers illegally.

Figure 3.1 shows that the increase in firm informality was widespread. Each dot represents a six-digit sector, the horizontal axes measures the share of informal firms (legal or non-compliant) in the total number of firms in that sector in 1998, and the vertical one the same share in 2018. As it turns out, 524 out of the 629 dots are above the 45-degree line, indicating that in most sectors the share of informal firms increased. Further, the color of the dots shows that this occurred in 208 out of 253 sectors in manufacturing, 106 out of 136 in commerce, and 210 out of 240 in services.

Figure 3.1: Firm informality at the six-digit sector level



Notes: authors' calculations with Census data.

Table 3.2 synthesizes information on employment, capital and value added; in each case, shares add to 100%. Resources moved in opposite directions between 1998 and 2018: employment in informal firms increased (all in non-compliant ones) and capital decreased, implying that formal ones became more capital intensive. There was little change in the contribution of formal and informal firms to value added, but there was substantial change in its composition within formal firms: large ones increased their share from 61 to 66% while the share of the rest fell.

Table 3.2: Resources and value added (Shares)

	1998	**2018**
Employment in:		
formal firms	67.6	61.58
informal and legal firms	21.82	20.77
informal and non-compliant firms	10.58	17.65
Capital in:		
formal firms	80.43	85.62
informal and legal firms	8.73	5.85
informal and non-compliant firms	10.83	8.53
Value added in:		
formal firms	84.58	85.66
informal and legal firms	6.97	5.34
informal and non-compliant firms	8.45	9

Notes: authors' calculations with Census data.

To consider changes at the six-digit level, we repeat the exercise shown in Figure 3.1 and find that the share of employment in informal firms increased in 473 out of the 629 sectors, the share of capital in 327, and the share of value added in 409. In other words, in the majority of sectors resources shifted towards informal firms.

Table 3.3 provides information on market shares, with the market defined as the gross value of domestic and export sales. The aggregate market share captured by formal firms increased marginally; a result due to manufacturing, as it fell in services and commerce.[13] At the six-digit level, the market share of informal firms increased in 408 out of the 629 sectors.

Table 3.3: Market shares

	Formal		**Informal**		**Total**	
	1998	**2018**	**1998**	**2018**	**1998**	**2018**
Manufacturing	78.9	87.4	21.1	12.6	100	100
Commerce	74.4	70.8	25.6	29.3	100	100
Services	75.4	69	24.6	31	100	100
Total	76.3	77.5	23.7	22.5	100	100

Notes: authors' calculations with Census data.

Summing up: between 1998 and 2018 there were contradictory changes in resource allocation, value added and market shares. In the aggregate, the share of informal firms increased as did the share of employment in those firms, while the share of capital fell. In parallel, the market

[13]The increase in the market share of formal firms in manufacturing is probably explained by the growth in exports. However, the 1998 census does not separate domestic from export sales so we cannot verify this. That said, recall that substantial economic activity is excluded from the census, mostly by informal firms in mobile premises. There is no data to measure their sales, but most likely the market share captured by informal firms exceeded 23%.

share of informal firms in manufactures fell, increased in services and commerce, and was practically constant in the aggregate. At the six-digit level changes were also heterogenous but in most sectors informality increased as measured by the share of firms, employment, capital, market share and value added. Within formal firms, large ones became more capital-intensive and produced a larger share of value added. Within informal ones, non-compliance increased. Altogether, these results indicate that between 1998 and 2018 a small number of large formal firms absorbed a larger share of capital and generated an increasing share of value added. A substantially larger number of small firms, mostly informal, absorbed more labor and produced a smaller share of value added.

3.5 Firm productivity by size, sector and formality status

3.5.1 Measurement of firm productivity

We follow Levinsohn and Petrin (2003) to measure firm productivity, applying the correction for functional dependence developed by Ackerberg et al. (2015). Consider the model:

$$VA_{ijt} = c_j + \mu_j L_{ijt} + \beta_j K_{ijt} + \Omega_{ijt} + e_{ijt} \tag{3.1}$$

where VA_{ijt} stands for value added of firm i in sector j at time period t, L_{ijt} for labor, K_{ijt} for capital, Ω_{ijt} for technical efficiency observed by the firm (but not by the econometrician) and e_{ijt} is a normally distributed error term (all variables in logs). We assume that L_{ijt} is chosen in period t but K_{ijt} in $t-1$, and that Ω_{ijt} follows the Markov process:

$$\Omega_{ijt} = g(\Omega_{ijt-1}) + u_{ijt} \tag{3.2}$$

We use intermediate inputs m_{ijt} as proxy for technical efficiency Ω_{ijt}. In parallel, we assume that current intermediate inputs are a function of current technical efficiency, capital and labor, and are adjusted immediately after an efficiency shock u_{ijt} is realized, so:

$$m_{ijt} = m_t(\Omega_{ijt}, L_{ijt}, K_{ijt}) \tag{3.3}$$

Where $m_t(\Omega_{ijt}, L_{ijt}, K_{ijt})$ is strictly increasing in Ω_{ijt}. Inverting the function $m_t(\Omega_{ijt}, L_{ijt}, K_{ijt})$ and denoting $\pi(.) = m^{-1}(.)$, equation (3.1) now becomes:

$$VA_{ijt} = c_j + \mu_j L_{ijt} + \beta_j K_{ijt} + \pi_t(\Omega_{ijt}, L_{ijt}, K_{ijt}) + v_{ijt} \tag{3.4}$$

Following Ackerberg et al. (2015), all coefficients in (3.4) are estimated simultaneously.

We drop all firms with zero capital, labor, or negative value added and use the STATA code written by Rovigatti and Mollisi (2020) to estimate these regressions with data from the 1998, 2003, 2008 and 2018 censuses.[14] We interpret the estimated values of β_j and μ_j as the structural parameters of each sector's production function. The estimation does not assume that $\beta_j + \mu_j$

[14] All coefficients are significant at the 95% confidence level; the tables with the detailed results are available from the authors.

= 1, so returns can vary across sectors. Because we only have price indices for value added at the three-digit level, we assume that β_j and μ_j apply to all firms in that sector.

With the estimated values of β_j and μ_j, we compute the (log) productivity of firm i in sector j as:

$$P_{ijt} = VA_{ijt} - \mu_j L_{ijt} - \beta_j K_{ijt} \tag{3.5}$$

Finally, note that (3.5) is a revenue-based measure of productivity, reflecting the firm's technical efficiency and the price received for its output. Clearly, when there is monopoly power, (3.5) will overstate productivity; a situation that may happen with a few large firms in services and commerce. Despite this possible bias, (3.5) is our preferred measure because it allows to compare firm productivity across and not only within-sectors and does not require assumptions about the elasticity of firms' demand functions or constancy of returns to scale. In any event, to test the robustness of our measure, we also computed measures of firms' physical and revenue productivity following Hsieh and Klenow (2009), denoted $TFPQ$ and $TFPR$. We find that in 1998 the correlation between (3.5) and $TFPQ$ and $TFPR$ was 0.95 and 0.86, respectively; and in 2018, 0.95 and 0.87.

3.5.2 Formal-informal productivity differences by size and sector

Table 3.4 calculates the differences in mean productivity between formal and informal firms in 1998 and 2018, separating them by sector and size. These are obtained as the coefficients of an OLS regression where formal firms are the omitted variable, and where we control by 3-digit sectors. [15]

The message from Table 3.4 is clear: regardless of how firms are classified, on average formal ones are more productive. Note that differences diminish with size and that, considering all firms, the difference in average productivity increased from 128% in 1998 to 139% in 2018.

3.5.3 Productivity distributions

Figure 3.2 presents the distribution of P_i in 1998 and 2018.[16] In both years, the median of the formal distribution is higher than the informal. The median of the complete distributions in 2018 is 7% higher than in 1998.

While the median and the mean of the formal productivity distributions in both years are higher than those of the informal distributions, there is considerable overlap between them. This implies that some informal firms have higher productivity than some formal ones; in fact, as we show below, some informal firms are very productive. The point here is that Mexico's informal sector is very heterogeneous, and that some firms may be informal not to avoid the

[15] The coefficients result from the OLS regression $P_i = \alpha + \beta D_i + \gamma_s + \varepsilon_i$, where $D_i = 1$ is informal and $D_i = 0$ otherwise and γ_s are controls for 3-digit sectors. The regression is equivalent to a mean test of productivity differences between formal and informal firms. All coefficients are statistically significant at the 95% confidence level.

[16] These are the distributions of P_i of firms in all sectors. Levy (2018) constructs similar distributions for 1998 and 2013 from the envelope of the 6-digit sector distributions of $TFPR_i$. The moments of those distributions are very similar to the ones in Figure 3.2. The result that the formal and informal productivity distributions overlap, but that the mean and median of the formal distribution is to the right of the informal one is robust, as is the result that overtime measures of dispersion increased (see below).

Table 3.4: Average productivity gap of informal firms relative to formal ones

	By sector			
	All	**Manufacturing**	**Services**	**Commerce**
1998				
Informal	(-) 1.282	(-) 1.461	(-) 1.090	(-) 1.386
s.e.	[0.0020]	[0.0048]	[0.0030]	[0.0031]
Obs.	2,546,761	317,879	867,590	1,361,292
R^2	0.201	0.28	0.201	0.177
2018				
Informal	(-) 1.394	(-) 1.485	(-) 1.233	(-) 1.508
s.e.	[0.0019]	[0.0057]	[0.0030]	[0.0028]
Obs.	4,045,080	403,573	1,621,465	2,020,042
R^2	0.208	0.181	0.151	0.244

	By size			
	1 – 5	**6 – 10**	**11 – 50**	**51+**
1998				
Informal	(-) 1.087	(-) 0.698	(-) 0.681	(-) 0.515
s.e.	[0.0024]	[0.0071]	[0.0109]	[0.0348]
Obs.	2,313,982	116,286	92,151	24,342
R^2	0.137	0.233	0.215	0.395
2018				
Informal	(-) 1.134	(-) 0.586	(-) 0.591	(-) 0.254
s.e.	[0.0027]	[0.0046]	[0.0068]	[0.0219]
Obs.	3,600,639	245,315	157,920	41,206
R^2	0.126	0.37	0.274	0.415

Notes: authors' calculations with Census data. Coefficients correspond to β in the model: $P_i = \alpha + \beta D_i + \gamma_s + \varepsilon_i$.

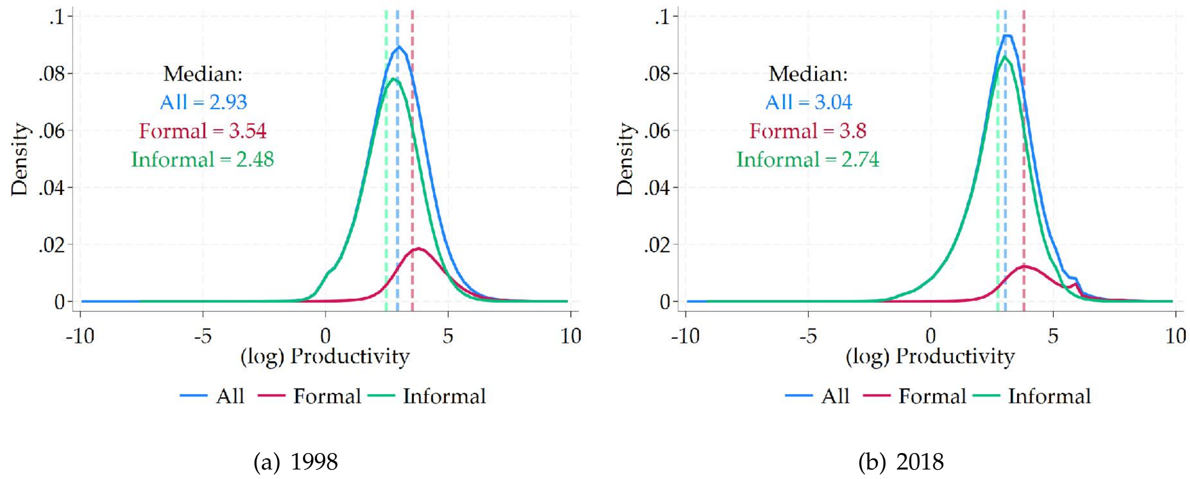
Law, but because they consider that non-salaried contracts with their workers are the best fit for their business model.

Panel (a) of Figure 3.3 overlaps the 1998 and 2018 distributions and panels (b) and (c) magnify the tails, below log one productivity and above log five. Between 1998 and 2018 the mean increased from 2.90 to 2.96 (or 6%), the standard deviation from 1.25 to 1.34, and the difference between firms in the $90^{th}/10^{th}$ percentiles from 3.14 to 3.26. Clearly, the productivity distribution polarized.

Panels (b) and (c) show that polarization resulted from a fattening of both tails, as the share of firms in each increased. Note that while the left-tail is almost wholly populated by informal firms, the right one is populated by a mix of both, and in fact in 2018 almost half of Mexico's high productivity firms were informal. Note that mean productivity fell in the left tail while it increased in the right one, again highlighting the polarization of the productivity distribution.

Figure 3.3 provides an initial insight to understand why TFP fell in Mexico between 1998 and 2018. On one hand, the number of high productivity firms doubled. A few of these survived since 1998 but, as we show below, the majority were new entrants. Regardless, these were the expected results from the measures to improve resource allocation. On the other, despite

Figure 3.2: Formal-informal firm productivity distributions



Notes: authors' calculations with Census data.

these measures, the number and share of low productivity firms also increased, and their average productivity fell. The balance yields a 6% increase in the mean between 2018 and 1998 and implies an annual growth rate of the simple average firm productivity of 0.3%; quite unimpressive, but at least positive. However, this result ignores changes in resource allocation among firms and, unfortunately, when this is considered as we do in the next sections, it is reversed.

3.6 Dynamic Olley-Pakes TFP decomposition: formal versus informal firms

3.6.1 Decomposition

We begin writing the expression for TFP subject to analysis. Let:

$$r_{ij} = K_{ij}^{\beta_j} \cdot L_{ij}^{\mu_j} \tag{3.6}$$

$$R = \sum_j \sum_i r_{ij} \tag{3.7}$$

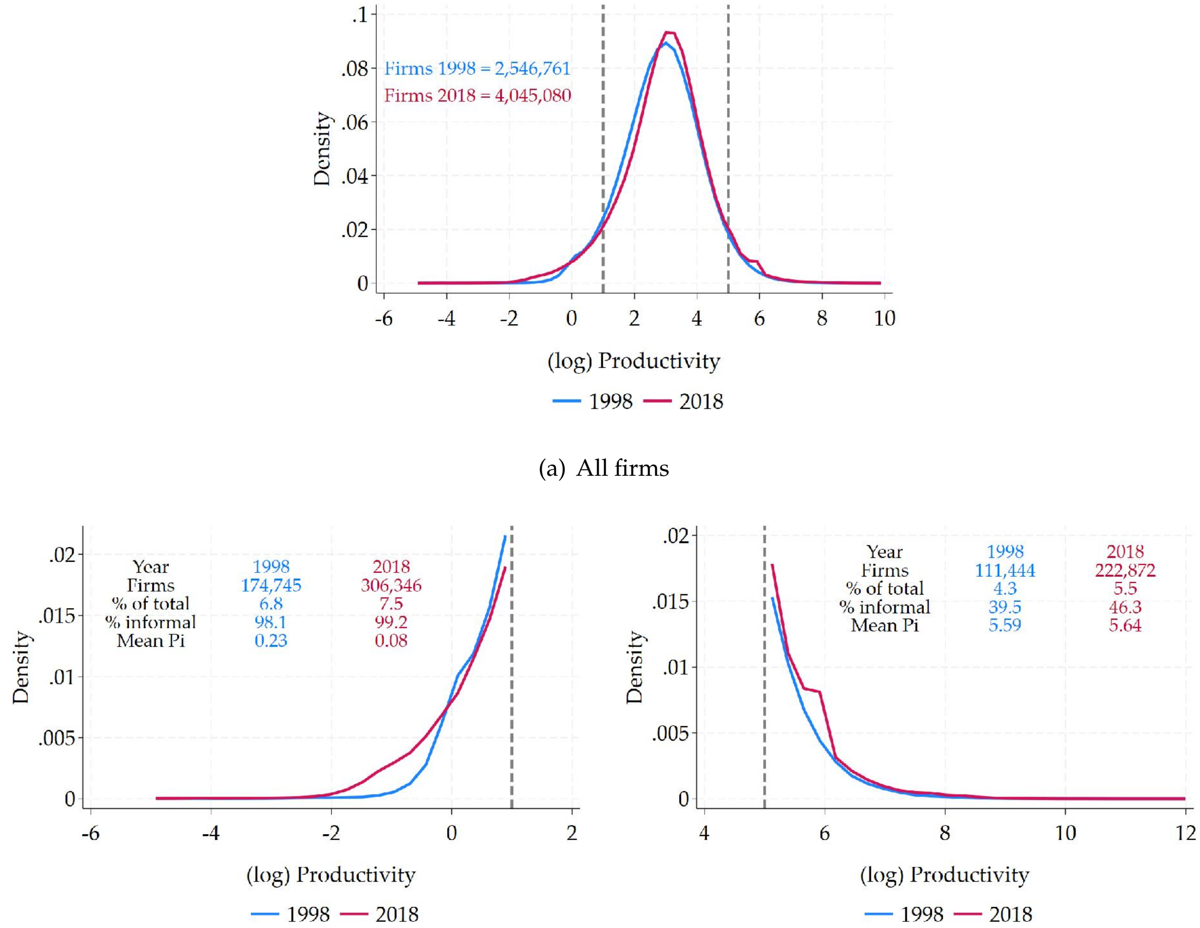
$$v_i = \frac{r_i}{R}; \qquad \sum_i v_i = 1 \tag{3.8}$$

so that r_{ij} are the resources captured by firm i in sector j, R is total resources, and v_i is the resource share corresponding to the i^{th} firm. TFP is the weighted average of firm productivity P_i, where the weights are the share of resources captured by each, v_i:

$$TFP \equiv \sum_i v_i \cdot P_i \tag{3.9}$$

Expression (3.9) serves to make two points: first, TFP depends on the joint distribution of P_i

Figure 3.3: 1998 and 2018 productivity distributions with amplified tails



(a) All firms

(b) Low productivity firms

(c) High productivity firms

Notes: authors' calculations with Census data.

and v_i. Second, it is additively decomposable, so one can compute TFP adding subsets of firms with their respective factor shares classified with different criteria.

Melitz and Polanec (2015), henceforth M-P, develop a methodology, labelled the dynamic Olley-Pakes productivity decomposition, to identify the contribution of firm exit, survival, and entry to the change in TFP between two periods, denoted here 1 and 2 (rather than 1998 and 2018, to simplify notation). Let X, S and E denote the set of exiting, surviving, and entering firms. Further, let $n_1(= n_S + n_X)$ and $n_2(= n_S + n_E)$ be the number of firms in the first and second period. TFP in each period is then:

$$TFP_1 = v_{S1} \cdot P_{S1} + v_X \cdot P_X \tag{3.10}$$

$$TFP_2 = v_{S2} \cdot P_{S2} + v_E \cdot P_E \tag{3.11}$$

where $v_{S1} = \sum_{n_S} v_{i1}$ is the share of resources captured by surviving firms in period 1, $P_{S1} = \sum_{n_S}(v_{i1}/v_{S1}) \cdot P_{i1}$ their weighted productivity, $v_X = \sum_{n_X} v_i$ the share of resources in exiting firms, $P_X = \sum_{n_X}(v_i/v_X) \cdot P_i$ their weighted productivity, and $v_{S1} + v_X = 1$. Similar expressions apply for period 2, except that entering firms replace exiting firms, and $v_{S2} + v_E = 1$.

M-P show that:

$$\Delta TFP = TFP_2 - TFP_1 = v_X \cdot (P_{S1} - P_X) + (P_{S2} - P_{S1}) + v_E \cdot (P_E - P_{S2}) \tag{3.12}$$

This is a very intuitive expression. The first term in the RHS measures the contribution of exiting firms to ΔTFP. It is positive if they are less productive than surviving firms in the first period, with the magnitude of the effect depending on the share of resources released by exiting firms, v_X. The second term captures the contribution of surviving firms and is positive if their weighted productivity increases. The last term measures the contribution of entering firms: they increase TFP if they are on average more productive than surviving ones in the second period, with the magnitude of the effect depending on the share of resources captured by them, v_E.

Expression (3.12) has a standard Schumpeterian interpretation: TFP increases if low productivity firms (relative to those that survive) die, if those that survive improve their performance, and if the ones that enter are more productive than the ones that survived. If all three conditions hold, there is Schumpeterian "creative destruction", and TFP unambiguously increases. If some do not, the net effect depends on the magnitude of each.

Expression (3.12) can be extended to separate between formal and informal entering and exiting firms, with F and I denoting each (where, quite naturally, $v_{XF} + v_{XI} + v_{S1} = 1$ and $v_{EF} + v_{EI} + v_{S2} = 1$):

$$\begin{aligned}\Delta TFP = [v_{XF} \cdot (P_{S1} - P_{XF}) + v_{XI} \cdot (P_{S1} - P_{XI})] + (P_{S2} - P_{S1}) +& \\ [v_{EF} \cdot (P_{EF} - P_{S2}) + v_{EI} \cdot (P_{EI} - P_{S2})]&\end{aligned} \tag{3.13}$$

To separate the term for surviving firms between formal and informal ones, note that firms can change status between periods: formal ones may remain formal (denoted here FF) or may turn informal (denoted FI); and similarly, informal firms may formalize, IF, or may remain in formal, II. Note as well that $v_{FF1} + v_{FI1} + v_{IF1} + v_{II1} = v_{S1}$ and similarly for period 2. Letting factor shares within surviving firms be $b_{FF1} = v_{FF1}/v_{S1}$ and so on (so that $b_{FF1} + b_{FI1} + b_{IF1} + b_{II1} = 1$ and similarly for period 2), we have:

$$\begin{aligned}P_{S2} - P_{S1} = (b_{FF2} \cdot P_{SFF2} + b_{FF1} \cdot P_{SFF1}) + (b_{FI2} \cdot P_{SFI2} + b_{FI1} \cdot P_{SFI1}) +& \\ (b_{IF2} \cdot P_{SIF2} + b_{IF1} \cdot P_{SIF1}) + (b_{II2} \cdot P_{SII2} + b_{II1} \cdot P_{SII1})&\end{aligned} \tag{3.14}$$

where the P's on the RHS of (3.14) are the weighted average of the productivity of each type of surviving firm in each period, where the weights are the factor shares captured by each. Substituting (3.14) in (3.13) we obtain the formula used in our calculations.

3.6.2 Panel of firms, firm productivity, and factor shares

Table 3.5 displays firm exit, survival, and entry between 1998 and 2018. The first row shows that there were 439,521 formal firms in 1998, of which 343,389 exited before 2018. Of the remaining 96,132 firms that were formal in 1998 and survived to 2018, 58,280 continued as formal (FF) and 37,852 changed their status to informal (FI). In parallel, 424,208 formal firms entered. Considering these, plus the formal ones that survived as formal, together with the 19,539 informal firms that survived but formalized (from the second row), yields a total of 502,027 formal firms in 2018. The second row is read similarly.

We highlight two facts in Table 3.5: first, 82% of the firms present in the market in 1998 exited before 2018, and 88% of those present in 2018 entered after 1998. Differently put, there was a lot of firm churning. That said, these figures underestimate churning because firms that entered after 1998 but exited before 2018 are not considered. In fact, using the data from the 2003, 2008 and 2013 censuses, it turns out that between 1998 and 2018 5.4 million firms exited and 6.9 million entered (an average of 285,000 and 364,000 per year, respectively). But even these figures underestimate churning because firms that enter and exit between two contiguous censuses are excluded (say, one that entered in 2005 but exited in 2007), and because the census only captures firms in urban areas in fixed premises. The point here is that firm churning in Mexico is substantially larger than what Table 3.5 suggests.

Table 3.5: Firm dynamics by formality status

	Starts 1998	Exit	Survival Total	Change type		Entry	Ends 2018
Formal	439,521	343,389	96,132	FF	58,280	424,208	502,027
				FI	37,852		
Informal	2,107,240	1,737,305	369,935	IF	19,539	3,154,805	3,543,053
				II	350,396		
Total	2,546,761	2,080,694	466,067			3,579,013	4,045,080

Notes: authors' calculations with Census data.

The second fact is that among surviving firms, 12% changed status, the majority towards informal. Among those that were formal, 39% survived and informalized, while among those that were informal, only 5% formalized. For every firm that changed from informal to formal status, almost two changed in the opposite direction. Differently put, the idea that informal firms that survive formalize is not supported in the Mexican data, even over a 20-year period.

3.6.3 Productivity decompositions

Table 3.6 shows factor shares and the weighted productivity of formal and informal exiting, surviving and entering firms. Substituting these values in equations (3.13) and (3.14) we obtain a key result: between 1998 and 2018 TFP fell by 7.4%, implying an annual growth rate of (-) 0.3%.[17]

[17] As a check on our results, we computed the change in aggregate TFP calculating the Solow residual from an aggregate production function using national accounts data. Setting the index of TFP at 1.00 in 1998, its value in 2018 was 0.899 (a fall of 9.1%). This can be contrasted with our findings using the O-P decomposition where, again setting the index of TFP at 1.00 in 1998, results in a value of 0.936 in 2018 (a fall of 7.4%). These results are very

What explains this dismal performance? Begin with exit. Since $v_{XF}(P_{S1} - P_{XF}) = 0.095$, the exit of formal firms contributed to increase TFP; and since $v_{XI}(P_{S1} - P_{XI}) = 0.262$, so did the exit of informal ones, in fact, significantly more. Altogether, exit by itself would have contributed to raise TFP by 35.7%, clearly a good outcome. That said, note that some exiting formal firms had higher productivity than some surviving informal ones (i.e., $P_{XF} > P_{SI}$, and by a large margin). If those exiting formal firms had survived, and those informal surviving ones exited, the exit process could have made a larger contribution to raise TFP. So, while exit helped, it was still problematic because relatively productive firms exited.

Table 3.6: Factor shares, firm productivity by formality status and ΔTFP

		Factor shares 1998	2018	**Weighted firm productivity** 1998	2018	**Contrib. to TFP**
Exit	Formal	$v_{XF} = 0.510$		$P_{XF} = 4.449$		0.095
	Informal	$v_{XI} = 0.160$		$P_{XI} = 2.998$		0.262
	All	$v_X = 0.671$		$P_X = 4.103$		0.357
Surv.	FF	$v_{SFF1} = 0.270$	$v_{SFF2} = 0.172$	$P_{SFF1} = 4.837$	$P_{SFF2} = 4.786$	0.027
	FI	$v_{SFI1} = 0.024$	$v_{SFI2} = 0.011$	$P_{SFI1} = 4.379$	$P_{SFI2} = 4.204$	(-) 0.097
	IF	$v_{SIF1} = 0.008$	$v_{SIF2} = 0.007$	$P_{SIF1} = 4.269$	$P_{SIF2} = 4.304$	0.038
	II	$v_{SII1} = 0.026$	$v_{SII2} = 0.016$	$P_{SII1} = 2.913$	$P_{SII2} = 2.876$	(-) 0.012
	All	$v_{S1} = 0.328$	$v_{S2} = 0.206$	$P_{S1} = 4.634$	$P_{S2} = 4.590$	(-) 0.044
Entry	Formal		$v_{EF} = 0.570$		$P_{EF} = 4.496$	(-) 0.054
	Informal		$v_{EI} = 0.223$		$P_{EI} = 3.096$	(-) 0.333
	All		$v_E = 0.793$		$P_E = 4.102$	(-) 0.387
Total		1.000	1.000	4.273	4.199	(-) 0.074

Notes: authors' calculations with Census data.

Survival is more problematic. Its contribution to TFP was negative because the weighted productivity of survivors fell ($P_{S2} < P_{S1}$). This fall reflects asymmetric behavior across the four firm statuses and is discussed in more detail in the next section but, all in all, surviving firms reduced TFP by 4.4%.

Entry is the most problematic. While formal entrants were more productive than informal ones ($P_{EF} > P_{EI}$) and attracted more resources ($v_{EF} > v_{EI}$), they were less productive than survivors ($P_{EF} < P_{S2}$); as a result, their contribution to ΔTFP was negative, (-) 5.4%. The same occurred with informal entrants, and by a much larger margin, (-) 33.3%. If the resources channeled to entrants had instead been allocated to survivors, TFP would have increased. In other words, it would have been better if new investments and new hirings had been allocated to expand existing firms rather than to create new ones, particularly informal ones, and the fact that this did not happen speaks volumes to the obstacles that Mexican firms face to grow.[18]

Very poor selection of entrants was the single most important factor behind the fall in TFP, reducing it by 38.7%, substantially larger than the negative contribution of survival (as noted,

close. The slightly larger fall in the first case is probably due to the fact that the whole economy is more informal than the segment captured in the Census.

[18]Consider three examples, each linked to the institutions discussed in section 3.2. First, if a firm grows it may need to change its contractual structure from non-salaried to salaried (for instance, to coordinate tasks among a larger set of workers). This, however, would increase substantially its flow and contingent costs of labor. Second, if firm growth implies crossing the threshold established in the tax code to qualify for the preferential regime, its after-tax profits can fall. And third, to issue bonds or attract new shareholders to increase its capital, the firm needs to be registered as a corporation and investors need to trust that their rights will be respected, a dubious proposition in a context of imperfect contract enforcement, particularly when it comes to small firms. That said, there may be other factors affecting firm growth, particularly of medium and large ones, like uncertain access to energy or costly finance.

4.4%). Selection at entry matters a lot because, as Table 3.5 shows, the vast majority of firms in 2018 entered after 1998 and because, as Table 3.6 shows, by 2018 surviving firms only captured 20.6% of all resources (v_{S2}) while entrants captured 79.3% (v_E). In other words, over the medium term, 20-years in this case, entry is key for TFP, and the fact that a lot of informal firms entered with lower productivity than survivors punished TFP considerably. Very poor selection at entry explains the fattening of the left-tail of the productivity distribution between 1998 and 2018 shown in Figure 3.3.

Further, note that even though informal entrants captured less than half of the resources than formal entrants did ($v_{EI} < 0.5 \cdot v_{EF}$), they captured more than all survivors ($v_{EI} > v_{S2}$). Moreover, their productivity was almost one-third lower than that of formal entrants ($P_{EI} \approx 0.68 \cdot P_{EF}$). In other words, informal entry mattered a lot. There is a lesson here: because most informal firms are very small, and each one captures a practically insignificant share of the economy's resources, they are usually thought of as a second-order issue, at least from the point of view of TFP. But this thinking is flawed because when added up these firms absorb a lot of resources (22% in 2018!), and because their productivity is very low, pulling the economy-wide average down.

The last column in Table 3.6 also allows to identify the contribution of formal and informal firms to the change in TFP. Altogether, the exit, entry and survival of formal firms, including those that formalized, increased TFP by 10.6%. In parallel, the exit, entry and survival of informal firms, including those that informalized, reduced TFP by 18%. Netting them out results in the 7.4% fall already noted. Clearly, the persistence of informal firms during this time period was extremely damaging to TFP in Mexico.

One more result. The capital stock of the firms considered in Tables 3.5 and 3.6 increased by 100% between 1998 and 2018 and the labor force by 85%, so that aggregate K/L increased. Nonetheless, TFP fell. Thus, contrary to what is at times stated, higher capital intensity does not always translate into more productivity. In Mexico's case, the increase in aggregate capital intensity hides considerable differences between formal and informal firms. Surviving formal firms became more capital intensive while surviving informal ones less; and entering formal firms were three times more capital intensive than entering informal ones. In other words, the formal sector became more capital intensive and the informal one less, but the weight of the latter dominated from the perspective of TFP.

3.7 Firm size, firm growth, and productivity

3.7.1 Firm growth and productivity

A significant advantage of our twenty-year panel is that, by focusing on survivors, we can observe the same firm in two time periods time and study the relation between firm growth and productivity. Table 3.7 provides the relevant data. Altogether, surviving firms grew 16.5%, from 8 to 9.3 workers over the 20-year period considered here, but their productivity fell by 4.4%. This result highlights the disconnect between changes in firm size and changes in productivity that occurs in a context of large misallocation and is the product of different behavior depending on firms' transitions.

Firms that remained formal (FF) grew 20%, from 45.8 to 55.1 workers but their productivity fell, although by 5.1% over 20 years. Despite the fall in their productivity, these firms made a positive contribution to ΔTFP, as can be seen in the last column of Table 3.6. The reason is that, within the set of surviving firms, they were the highest productivity ones and they attracted resources from other firms with lower productivity ($b_{SFF2} > b_{SFF1}$). In this case, resource reallocation within survivors compensated the disappointing productivity performance of firms that survived as formal.

On the other hand, firms that informalized (FI) shrunk and their productivity fell by 17.5%; clearly, it would have been better if they had died. Firms that formalized (IF) increased their size considerably, by 39%, and their productivity, although again by a small amount, 3.5% over 20 years. Finally, those that stayed informal (II) grew 5% but their productivity fell by 3.7%; again, it would have been better if they had died.

Table 3.7: Average firm size and (log) productivity of surviving firms

		Average size			Weighted (log) productivity		
	Number	**1998**	**2018**	**% change**	**1998**	**2018**	**% change**
FF	58, 280	45.8	55.1	20.3	4.837	4.786	(-)5.1
FI	37,852	7.6	6.1	(-)19.8	4.379	4.204	(-)17.5
IF	19,539	6.7	9.3	38.8	4.269	4.304	3.5
II	350,396	1.8	1.9	5.5	2.913	2.876	(-)3.7
All	466,067	8	9.3	16.2	4.634	4.59	(-)4.4

Notes: authors' calculations with Census data.

Table 3.7 allows two observations. First, informality was a status that allowed firms that should have exited to survive. Even though firms that survived as informal attracted very few resources, their productivity was so low that they more than offset the modest contribution to ΔTFP from firms that survived as formal or formalized. And second, it is often stated that "informal firms that survive formalize, grow, and become more productive". Unfortunately, in the case of Mexico this statement applies only to 19,539 out of the 369,935 firms that were informal in 1998 and survived to 2018; the remaining 95% did not formalize or become more productive.

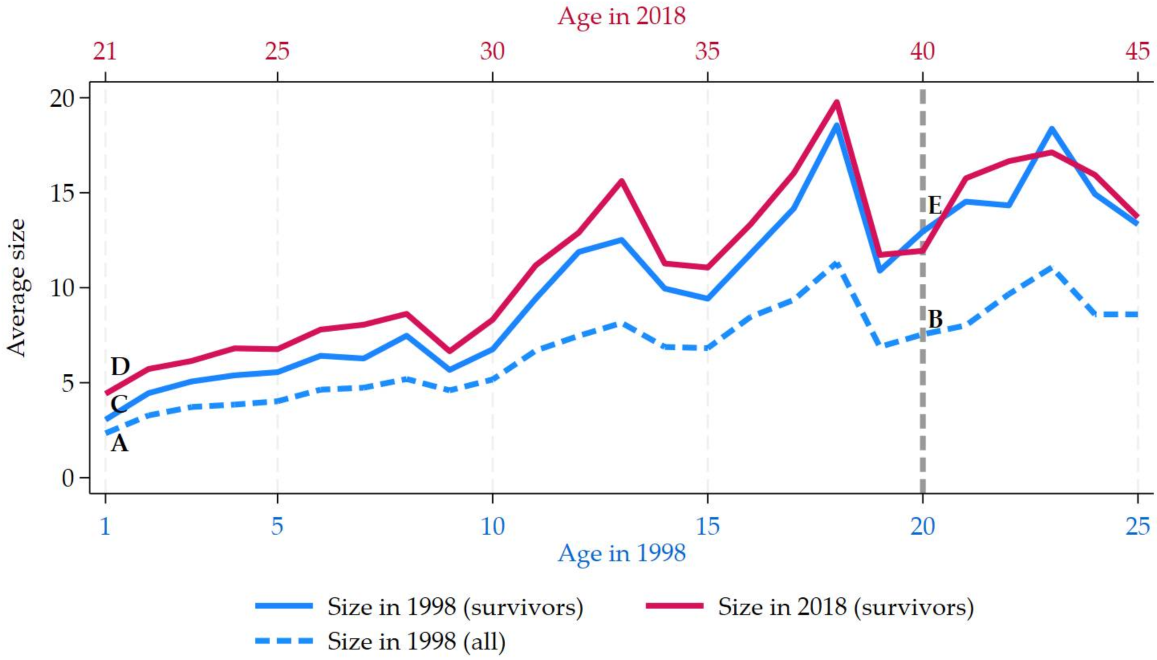
3.7.2 Age-size profiles and firm growth

Figure 3.4 shows the relation between the size and age of surviving firms in 1998 (blue line) and in 2018 (red line), and of all firms in 1998 (dotted blue line). The lower horizontal axes depicts firms' age in 1998 and the upper one the age of those that survived to 2018.

Studies that infer firm growth from firms' age-size profiles with cross-sectional data, as in Hsieh and Klenow (2014), focus on the dotted line and use data from all firms in that year. In this case, one year old firms in 1998 were at point A and had on average 2.34 workers and, 20 years later, would be at point B, with an average of 7.55 workers (point B), implying that they grew by 322%.

However, not all firms at point A survived 20 years and reached point B; in fact, most did not. To measure firm growth properly, we need panel data from two periods and focus on the same firms, that is, those that survive two decades, as captured by the solid blue and red lines. Critically, firm growth is the vertical movement between the blue and red lines, not the horizontal movement along the blue line. Point C represents one-year old firms in 1998,

Figure 3.4: Firms' age-size profiles



Notes: authors' calculations with Census data.

meaning they entered in 1997; on average, they had 3.05 workers. In 2018 they were 21 years old and were at point D with, on average, 4.42 workers, not at point E (12.96 workers). Thus, one-year old firms in 1998 firms grew by 26.3% in 2018, far from the 322% implied by the dotted line, or the 424% implied by moving along the blue solid line.

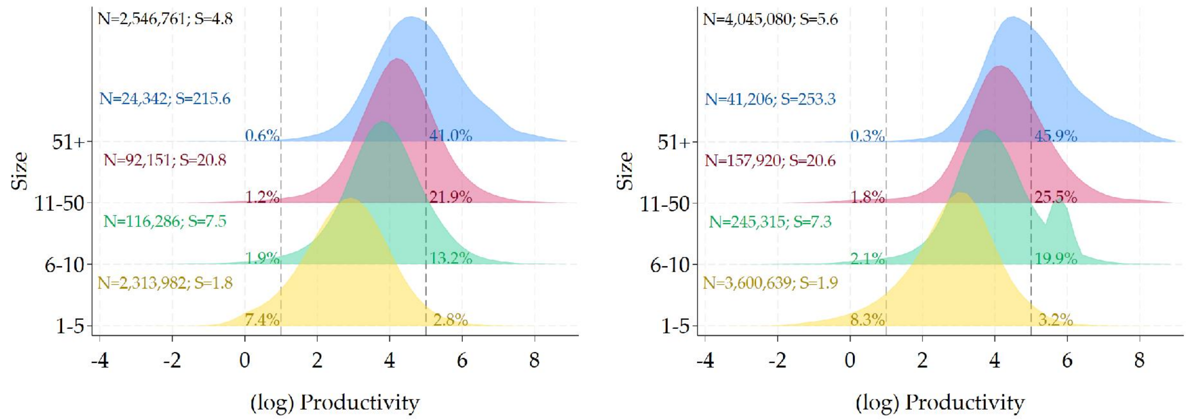
Because the vertical distance between the blue and red lines is fairly constant, firm growth is decreasing in size. Firms with approximately 5 workers in 1998 had 6.1 in 2018, so they grew 22%; firms with approximately 12 workers in 1998 had 13 in 2018, so they grew 8.3%; and so on. Importantly, firms with 15 or more workers in 1998 only grew 5%; an almost insignificant amount considering that this occurred over two decades (an annual growth rate of 0.2%, less than one tenth the average growth rate of GDP in that period!). The point here is that average firm growth was caused mostly by very small and small firms, whose growth in percentage terms is inevitable large, as they pass from 2 to 3 or 3 to 4 workers; medium and large ones hardly grew.

3.7.3 Changes in average firm size and changes in productivity

Figure 3.5 depicts the same productivity distributions presented in Figure 3.3 but segmented by ranges of size. For each size range, we show the share of firms in each tail, and to the left of the distributions, the number of firms in each range, N, their average size, S, with totals in black.

Various observations are relevant. First, recall from Table 3.5 that almost nine out of ten firms present in 2018 did not exist in 1998, so that differences between these two decades mostly reflect exit and entry, not survival. Second, note that considering all firms, average size increased from 4.8 to 5.6 workers, or almost 17% However, this was mostly accounted for by large firms:

Figure 3.5: Productivity distributions by range of size



(a) 1998 (b) 2018

Notes: authors' calculations with Census data.

253.3 workers in 2018 vs. 215.6 in 1998. Very small firms had almost the same number (1.9 vs. 1.8), and small and medium ones actually had fewer. In turn, recall that large surviving firms hardly grew, so that the increase in their average size is explained mostly by entry, not growth. In turn, this implies that practically all the increase in average firm size in these two decades resulted from the entry of large firms.[19]

Third, it is evident that large firms are on average more productive: in both years they have the smallest left tail and the largest right one. This supports the association commonly made between firm size and productivity. That said, it is also evident that the distributions by size range overlap; a fact that points out that there are some very small firms (and small and medium ones) with higher productivity than some large ones.[20]

Lastly, comparing the 1998 and 2018 distributions, note that the left tail contracted only for large firms and expanded for all others. This asymmetric behavior provides further insights into one of the main results of the Olley-Pakes decompositions presented before, namely, that entry was the single most important factor that depressed TFP between 1998 and 2018. The point here is that entry itself was heterogeneous: in the case of large firms, it increased size and TFP, but for the rest, it did neither.

Although for confidentiality reasons we cannot identify them individually, it is very likely that exporting firms are among the large high productive entrants, a result intimately associated with the trade liberalization efforts mentioned in the introduction. The 2018 census reports 11,387 firms that are direct exporters, who are on average 51 times larger than the average firm, 2.6 times more capital intensive, and pay 40% higher wages (Levy and Fentanes, 2022). Of these, 74% entered in or after 1994, when the North American Free Trade Agreement began.

[19] As discussed before, very small and small surviving firms grew, but substantially more entered with a lower size, so that their average size in 2018 was almost the same as in 1998: 1.9 vs. 1.8 for the case of very small firms, and 7.3 vs. 7.5 in the case of small ones.

[20] In fact, in absolute numbers there substantially more very small high productivity firms than large ones. For example, for 1998, 2.8% of 2,313,982 exceeds 41% of 24,342 (64,791 vs. 9,980).

3.8 Dynamic Olley-Pakes decomposition: manufacturing versus services and commerce

3.8.1 Firm dynamics and resource allocation

In this section we classify firms into manufacturing (denoted M) and services and commerce (R, for rest). We again use expressions (3.13) and (3.14) to decompose ΔTFP, simply substituting M for F and R for I. Table 3.8, analogous to Table 3.5, describes firm dynamics (except that in this case there are no changes of sector within survivors). By construction, the totals for exit, survival and entry are the same as in Table 3.5. A key point to note is that over this period the share of employment in manufacturing fell from 35.5 to 27.2%, and its share of capital from 45.4 to 40.5

Table 3.8: Firm dynamics by sector

	Starts 1998	**Exit**	**Survival**	**Entry**	**Ends 2018**
Manufactures	317,879	270,236	47,643	356,597	404,240
Rest	2,228,882	1,810,458	418,424	3,222,416	3,640,840
Total	2,546,761	2,080,694	466,067	3,579,013	4,045,080

Notes: authors' calculations with Census data.

3.8.2 Productivity decompositions

Table 3.9 presents factor shares, the weighted productivity of exiting, surviving, and entering firms and the contribution of each to ΔTFP. Four observations are of interest. First, productivity in manufacturing is higher than in services and commerce and, more importantly, the gap increased. Since the main difference between them is their exposure to international trade, it is difficult to avoid the conclusion that the trade liberalization measures are mainly responsible for the relatively better performance of manufactures. This conclusion is buttressed by the asymmetries in the behavior of productivity among surviving firms: in manufacturing it increased by 12.2% ($= P_{SM2} - P_{SM1}$), while in the other sectors it fell, by 20% ($= P_{SR2} - P_{SR1}$); differently put, services and commerce fully account for the productivity fall among surviving firms. It is also buttressed by the fact that in manufacturing, entering firms are more productive than exiting ones ($P_{EM} > P_{XM}$), while the opposite occurs in services and commerce ($P_{ER} < P_{XR}$). And, finally, it is buttressed by the fact that entering firms are substantially more productive in manufacturing than in services and commerce ($P_{EM} > P_{ER}$).

In other words, overtime manufacturing behaved differently, better than services and commerce. That said, note that entering manufacturing firms are less productive than surviving ones ($P_{EM} < P_{S2}$), a phenomenon due to the fact that entry of informal low productivity firms into manufactures was large.[21]

Second, while manufacturing performed better, its factor share fell from 27.6% in 1998 (= $v_{XM} + vS_{M1}$) to 26.1% in 2018 (= $v_{SM2} + v_{EM}$). This finding is clear evidence of misallocation across

[21] There were 317,879 manufacturing firms in 1998, 79,258 formal and 238,621 informal. Of these, 47,643 survived to 2018, 16,346 formal and 31,297 informal. Among surviving firms, 5,999 transited from formal to informal, and 2,216 in the opposite direction. In parallel, 46,921 formal and 309,676 informal firms entered, yielding a total of 404,240 firms in 2018, 59,484 formal and 344,756 informal. Note that among surviving firms more transited from formality into informality than vice versa; almost by a ratio of three to one.

Table 3.9: Factor shares, firm productivity by sector and ΔTFP

		Factor shares 1998	Factor shares 2018	Weighted firm productivity 1998	Weighted firm productivity 2018	Contribut. to TFP
Exit	M	$v_{XM} = 0.157$		$P_{XM} = 4.246$		0.061
	R	$v_{XR} = 0.513$		$P_{XR} = 4.056$		0.296
	All	$v_X = 0.670$		$P_X = 4.103$		0.357
Surv.	M	$v_{SM1} = 0.119$	$v_{SM2} = 0.088$	$P_{SM1} = 4.857$	$P_{SM2} = 4.979$	0.363
	R	$v_{SR1} = 0.209$	$v_{SR2} = 0.118$	$P_{SR1} = 4.507$	$P_{SR2} = 4.302$	(-) 0.407
	All	$v_{S1} = 0.329$	$v_{S2} = 0.206$	$P_{S1} = 4.634$	$P_{S2} = 4.590$	(-) 0.044
Entry	M		$v_{EM} = 0.173$		$P_{EM} = 4.530$	(-) 0.010
	R		$v_{ER} = 0.620$		$P_{ER} = 3.980$	(-) 0.378
	All		$v_E = 0.793$		$P_E = 4.102$	(-) 0.388
Total		1.000	1.000	4.273	4.199	(-) 0.074

Notes: authors' calculations with Census data.

sectors, as the high productivity sector of the economy contracted. It also implies that from the point of view of TFP, the performance of services and commerce is extremely relevant, and that when they underperform, they punish TFP considerably.[22]

Third, contrasting Tables 3.6 and 3.9, it is clear that regardless of whether we consider exit, survival, or entry, the differences in productivity between formal vs. informal firms are larger than those between firms in manufacturing vs. services and commerce, and in all cases by large margins. This observation is critical because it highlights that from the point of view of TFP, the contractual differences between firms matter substantially more than their differences in exposure to international trade. As we show below, TFP would increase much more in Mexico closing the productivity gap between formal and informal firms than by closing it between firms in manufactures versus the other two sectors.

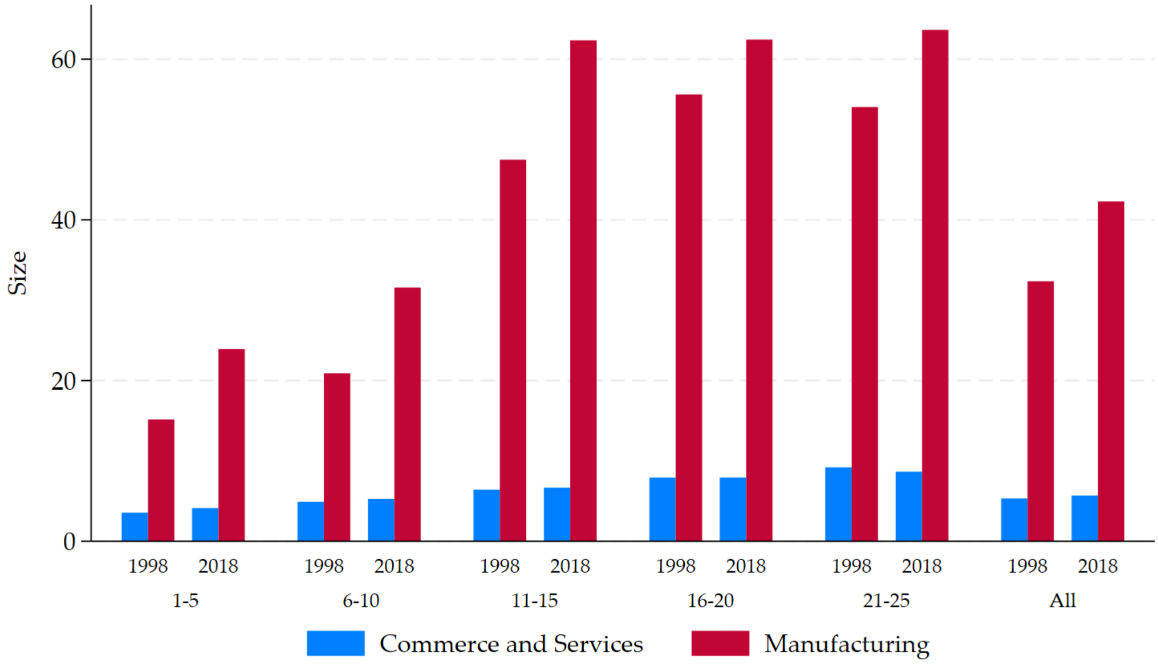
Finally, we describe the contribution of each sector to ΔTFP. Exit in manufacturing raised TFP by 6.1% while services and commerce by 29.6%. The exit of manufacturing firms contributed little because their productivity was higher ($P_{XM} > P_{XR}$), and because the resources involved were substantially smaller ($v_{XM} < v_{XR}$). In parallel, as noted, services and commerce fully account for the 4.4% fall in productivity among survivors. Finally, firms in services and commerce almost fully account for the negative contribution of entry to ΔTFP: (-) 37.8% vs. (-) 1% for manufacturing; a result due to the fact that their factor share was higher ($v_{ER} > v_{EM}$) and the gap vis-à-vis the productivity of surviving firms was larger, that is, in absolute values, $(P_{S2} - P_{ER}) > (P_{S2} - P_{EM})$. In the end, during the two decades considered here, manufacturing played a relatively modest role in the changes in TFP; services and commerce had the upper hand.

3.8.3 Firm growth

We close this section discussing firm growth from the sector perspective, focusing again on surviving firms. Note from Table 3.8 that out of the 466,067 surviving firms, only 47,643 are in manufacturing ($\approx 10\%$). With that observation, the horizontal axes in Figure 3.6 groups firms

[22] The share of manufactures in GDP fell from 18% in 1998 to 15.3% in 2018. Our results are consistent Rodrik's (2016) 'premature deindustrialization' hypothesis. What is notable in Mexico's case is that deindustrialization occurred despite the very successful performance of manufacturing exports.

Figure 3.6: Size of surviving firms: manufacturing vs. services and commerce



Notes: authors' calculations with Census data.

by sector and age in 1998 (between one and five years old, six and ten, and so on). The vertical axes shows their size, given by the number of workers.

Considering firms in all sectors, average size increased by 16%, from 8 to 9.3 workers (Table 3.7). However, Figure 3.6 shows that the differences between manufactures and services and commerce are dramatic and widened in these two decades. Average size in manufactures was 32.4 in 1998 and 42.3 in 2018, an increase of 30%; in contrast, in services and commerce it was 5.3 and 5.7, respectively, an increase of 7.5%. Differently put, the 16% increase in the size of surviving firms was basically driven by manufactures (despite the fact they represented only 10% of survivors).

Thus, Figure 3.6 provides further evidence that manufactures behaved differently than services and commerce. As shown before, it was the only sector to make a positive contribution to ΔTFP and, as shown here, firms grew substantially more. Unfortunately, as already noted, manufactures was unable to increase its share of resources and despite its relatively better performance, TFP fell.

3.9 Two back-of-the envelope calculations: formality vs. sector composition

What would have happened to TFP if between 1998 and 2018 the formal-informal segmentation of the economy had disappeared? Answering this question requires a model capturing the impact of the institutions alluded to in section 3.2 on firm and worker behavior. Clearly, changing them would impact occupational choices, the size distribution of firms, the dynamics of entry, survival, and exit, the patterns of firm growth, and incomes and the size of the market, among many variables. An exceedingly difficult task not attempted here.

Rather, in this section we carry out two mechanical exercises. In the first one we assume that between 1998 and 2018 the productivity of informal surviving firms converges to that of formal surviving firms, and the productivity of entering informal firms equals that of entering formal ones. The exercise is equivalent to a scenario where the market share of informal firms falls from 24% to 0%, so that formal ones make all investments, hire all workers, and produce all goods and services. In this scenario, between 1998 and 2018 TFP would have increased by 27%, for an annual growth rate of 1.2%.[23] This result compares to the annual growth rate of (-) 0.3% estimated in section 3.6 and provides another angle on the extent to which informal firms depress productivity growth.

In the second exercise we assume that between 1998 and 2018, the productivity of surviving and entering firms in services and commerce converges to that of manufacturing. In this case, TFP would have increased by 12.3%, less than half of the increase in the 'no informality' case. This result is explained by the fact that in this case TFP continues to be punished by the presence of informal firms and supports the following observation: from the point of view of TFP, the formal-informal composition of the economy matters substantially more than its sector composition.

3.10 Conclusions

In this paper we exploited a very rich and, by Latin American standards, unique firm database, to understand why, despite many reforms to increase efficiency and a boom in manufacturing exports, TFP fell in Mexico in the last decades. We have six results: first, between 1998 and 2018 firm informality increased in the aggregate and in most six-digit sectors, productivity differences between formal and informal firms widened, and the distribution of firm productivity polarized. In parallel, the market share of formal firms increased in manufacturing, fell in services and commerce, and increased marginally in the aggregate, from 76 to 77%.

Second, using a 20-year panel to study firm dynamics, we find large churning: eight out of ten firms present in 1998 exited before 2018 and nine out of ten in 2018 entered after 1998. However, this churning was useless as TFP fell 7.4%. Exit raised TFP because many unproductive informal firms exited, although troublingly some higher productivity formal firms also exited. Survival lowered TFP because, on balance, the productivity of surviving firms fell. Entry also lowered TFP because many informal low productivity firms entered; in fact, very poor selection at entry was the single most important factor punishing TFP. All in all, firm dynamics were dysfunctional.

Third, for each surviving informal firm that formalized, two surviving formal ones informalized. Only 5% of surviving informal firms followed the expected path of "growing, formalizing and becoming more productive"; the remaining 95% neither grew, nor formalized nor became more productive. Further, formal firms that survived by informalizing became less productive. Altogether, the relation between changes in size of surviving firms and changes in their productivity was the opposite of what was expected, as average size increased but productivity

[23]This growth rate –though still lower than that observed in many East Asian countries– would have exceeded that of Canada (0.26%) and the United States (0.66%) over the same period. In other words, rather than divergence, there would have been convergence in TFP between Mexico and its Nafta partners.

fell.

Fourth, the increase in average firm size between 1998 and 2018 was driven almost exclusively by the entry of relatively few large firms. Surviving medium and large firms hardly grew, particularly in services and commerce, and many small firms entered. As a result, the distribution of firm size also polarized.

Fifth, manufacturing behaved differently than services and commerce, as its TFP increased while it fell in the other two sectors. That said, its performance was far from stellar because despite the many measures promoted to increase efficiency, informal entry into manufacturing continued. Moreover, its contribution to aggregate TFP was diluted because, despite being the higher productivity sector, its share of resources fell.

Finally, we find that productivity differences between formal and informal firms are larger than those between firms in manufacturing and other sectors, implying that from the point of view of TFP, the formal-informal composition of the economy matters more than the sector composition.

Our findings are based on the dynamic productivity decomposition proposed by Melitz and Polanec (2015) to an economy with a large informal sector. As opposed to the "Solow residual" obtained from an aggregate production function, the O-P decomposition studies the path of TFP following the patterns of exit, survival and entry of individual firms. The Solow residual is usually thought of as a black box; "a measure of our ignorance". This contrasts with the O-P decomposition, where changes in TFP are derived from the performance of individual firms; in our case, over 6.5 million. The O-P decomposition sheds considerable light on the behavior of TFP in Mexico because it highlights the critical role played by resource misallocation across and within sectors.

When firms are classified by sector, the Olley-Pakes decomposition highlights the asymmetric behavior of manufactures versus services and commerce, and calls attention to the fact that while manufacturing TFP may increase, aggregate TFP can fall. While it is often the case the data limitations preclude analysis of productivity in services and commerce, this finding suggests caution when extrapolating the results of studies focused only on manufacturing. In contexts of large misallocation, manufacturing TFP may increase while its share of resources falls, and its positive contribution to aggregate TFP may be offset by the negative contribution of other sectors, as was the case in Mexico.

In parallel, when firms are classified by formality status, the Olley-Pakes decomposition highlights the fact the TFP can increase in the formal sector and fall in the informal one. Again, while it is often the case that data limitations preclude analysis of informal firms, this finding suggests caution when extrapolating the results of analysis of TFP that focus only on formal ones. Any individual informal firm is almost irrelevant; jointly they can make all the difference, as was also the case in Mexico.

Finally, the Olley-Pakes decomposition provides a useful complement to analyses of the impact of individual policies on TFP. Undoubtedly, the advantage of these analysis is that they carefully identify the impact of a single policy and the mechanisms through which it impacts TFP. However, by focusing on an individual tree, they miss the interaction with other trees, an extremely relevant consideration when other trees behave differently from the tree under study and may determine the fate of the forest. The point here is that to obtain a fuller understanding of the determinants of changes in aggregate TFP, we need both: studies of individual policies with techniques that allow to identify causality, and studies of how multiple policies interact and determine the overall outcome, even if one cannot identify the individual contribution of each, as this paper attempted.

Our findings have substantive implications for policy in Mexico and, we would argue, for countries with large informal sectors. First, they highlight that from the point of view of TFP, the formal-informal segmentation of the economy is very costly, and that this segmentation can persist and in fact increase even in the context of reforms like privatizations, creation of regulatory bodies to promote competition, and trade liberalization.

Second, they reflect the inconsistent nature of the policymaking process in Mexico. At the end of the day, the dysfunctional nature of its firm dynamics in the period studied here show that the efficiency-enhancing reforms promoted since 1990 to increase TFP could not counteract other forces in the economy operating in the opposite direction.[24]

Third, they highlight that, ignoring social protection issues, informality is a "market competition problem". Informal firms survive or are continuously created because they can adapt to shocks with more ease than formal ones; and because they are implicitly subsidized by the dual nature of the Mexico's social insurance architecture, and by special tax regimes. In parallel, formal firms have more difficulty responding to shocks, and are implicitly taxed by flaws in the social insurance regime and by enforcement of regulations proportional to firm size; as well as hindered by a weak contracting environment. Because formal and informal firms co-exist in most narrowly defined markets, the result is that competition is heavily distorted, weakening the connection between firm size, firm growth, and productivity. The point here is that this "market competition problem" could not be addressed by the privatization of state-owned enterprises and the trade liberalization measures promoted by Mexico, including its fourteen trade agreements; and was legally beyond the reach of the anti-trust authorities that were created in parallel.

Fourth, our findings suggest that in countries with large informal sectors, policymakers need to exercise care with policies that promote entrepreneurship. Entry of new firms should not be an objective by itself; what matters is that entrants be better than incumbents. Across-the board the board promotion of entry might result in the proliferation of informal firms that can end-up hurting productivity.

Fifth, our findings indicate that successful export performance, particularly in manufacturing, need not always be an 'engine of TFP growth'. This is not to say that manufacturing exports are not welcome; they are, and without them Mexico's productivity performance would have been even more dismal. But it is to say that they cannot offset the institutions and policies that generate the formal-informal divide. Mexico's experience is thus a cautionary tale not in the sense that countries with large informal sectors should not open to international trade, but in the sense that, in parallel, they need to do much more to fully reap its benefits. Differently put, trade reform or, for that matter, privatizations or anti-trust policies, are not a substitute for tackling the roots of the formal-informal divide, and the first without the second can result

[24] Levy (2018) documents that between 1998 and 2018, tax, labor and social insurance regulations changed, favoring informality, at the same time that the contracting environment faced by firms deteriorated.

in a situation where a segment of the economy performs very well, and the rest stagnates.

Sixth, policymakers often pay large attention to manufacturing, hoping that improving its performance will increase aggregate TFP: industrial or productive development policies, credit from development banks, subsidies for R&D, free trade areas, and so on. However, our findings underline the importance of focusing on services and commerce. These sectors can more than offset manufacturing's positive behavior, more so if their share of resources increases. Differently put, to increase aggregate TFP, policymakers need to pay attention to services and commerce, even if this is more challenging because informality in these sectors is more prevalent.

Bibliography

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Álvarez, F., Eslava, M., Sanguinetti, P., Toledo, M., Alves, G., Daude, C., and Allub, L. (2018). Red 2018. instituciones para la productividad: hacia un mejor entorno empresarial.

Bento, P. and Restuccia, D. (2017). Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics*, 9(3):267–303.

Buera, F. J. and Shin, Y. (2013). Financial frictions and the persistence of history: A quantitative exploration. *Journal of Political Economy*, 121(2):221–272.

Busso, M., Fazio, M., and Algazi, S. (2012). (in) formal and (un) productive: The productivity costs of excessive informality in mexico.

Busso, M., Fentanes, O., and Levy, S. (2018). The longitudinal linkage of mexico's economic census 1999-2014.

Eslava, M., Haltiwanger, J., and Pinzón, Á. (2022). Job creation in colombia versus the usa:'up-or-out dynamics' meet 'the life cycle of plants'. *Economica*, 89(355):511–539.

Fernández-Arias, E. and Fernández-Arias, N. (2021). The latin american growth shortfall: Productivity and inequality. *Background paper for the*.

Gasparini, L. and Tornarolli, L. (2009). Labor informality in latin america and the caribbean: Patterns and trends from household survey microdata. *Desarrollo y sociedad*, (63):13–80.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448.

Hsieh, C.-T. and Klenow, P. J. (2014). The life cycle of plants in india and mexico. *The Quarterly Journal of Economics*, 129(3):1035–1084.

La Porta, R. and Shleifer, A. (2014). Informality and development. *Journal of economic perspectives*, 28(3):109–126.

Lederman, D., Messina, J., Pienknagura, S., and Rigolini, J. (2014). *Latin American entrepreneurs: many firms but little innovation*. World Bank Publications.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The review of economic studies*, 70(2):317–341.

Levy, S. (2018). *Under-rewarded efforts: The elusive quest for prosperity in Mexico*. Inter-American Development Bank.

Levy, S. and Fentanes, O. (2022). Nafta-usmca and wages in mexico. *Brookings Institution*.

Maurizio, R. (2021). Employment and informality in latin america and the caribbean: an insufficient and unequal recovery. *ILO Labour Overview Series, Latin America and the Caribbean*.

Melitz, M. J. and Polanec, S. (2015). Dynamic olley-pakes productivity decomposition with entry and exit. *The Rand journal of economics*, 46(2):362–375.

Pagés-Serra, C. (2010). The age of productivity: Transforming economies from the bottom up. Technical report, Inter-American Development Bank.

Poschke, M. (2018). The firm size distribution across countries and skill-biased change in entrepreneurial technology. *American Economic Journal: Macroeconomics*, 10(3):1–41.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, 11(4):707–720.

Rodrik, D. (2016). Premature deindustrialization. *Journal of economic growth*, 21:1–33.

Rovigatti, G. and Mollisi, V. (2020). Prodest: Stata module for production function estimation based on the control function approach.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence from brazil. *American Economic Review*, 108(8):2015–2047.