Historical Persistence and the Dynamics of Development

Evidence from plant data

Oscar Fentanes*, Jonas Gathen†‡

23 Nov 2022

Abstract

How important is historical persistence in explaining observed economic growth? Drawing on 40 years of panel data on Indonesian manufacturing plants, we show that initial conditions in 1975, characterized by small plants and a missing right-tail, are highly predictive of future growth dynamics. To disentangle the role of initial conditions from other main drivers of plant dynamics and quantify its importance, we build a structural model of non-stationary plant dynamics that closely matches observed growth dynamics. We find that initial conditions account for 20% of observed growth, compared to technology's 60% - the main driver in standard models. Importantly, we find that initial conditions do not become less important over time. The reason is that initial conditions induce only a slow transition to a new stationary distribution due to the presence of sizable frictions, while changes in entry and aggregate technology continuously create new potential for growth.

*Toulouse School of Economics (TSE), Email: oscar.fentanes@tse-fr.eu

†Toulouse School of Economics (TSE), Email: jonas.gathen@tse-fr.eu.

‡We are grateful for many useful comments by Fabrice Collard, Christian Hellwig, Ben Moll, Yongseok Shin, Francisco Buera, Eugenia Gonzalez-Aguado, Miguel Zerecero, Andreas Schaab 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 participants at 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.

1 Introduction

Historical persistence is at the core of theories of economic growth and development. For example, in the neoclassical growth model, initial conditions in factor inputs and technology drive subsequent factor accumulation - a mechanism considered central for post World War 2 catch-up growth of Japan and European countries as well as rapid growth of a number of East Asian countries.[1] More recent theories on the misallocation of resources have emphasized an alternative growth narrative whereby changes in frictions drive growth by undoing misallocation of resources and improving the selection of firms. In these theories, the initial state of misallocation in the economy is central to explain future growth. For example, Buera and Shin (2013) show how an initially sufficiently misallocated economy could explain the growth experience of Asian miracle economies after market reforms. And Song, Storesletten, and Zilibotti (2011) explain China's decades of economic growth in large part through the initial misallocation of resources in state-owned firms and their slow subsequent reallocation towards more productive private sector firms.

In this paper, we provide a unified model and estimation framework that allows to study dynamic economies in which growth is influenced both by initial conditions and aggregate changes over time. The model framework allows initial conditions to play a flexible role both through the accumulation of factors of production and through the reallocation of resources across firms in the economy. In this model setup, initial conditions are given by the entire distribution of producing firms, requiring to track the distribution of firms over time. To study realistic features of dynamic economies and to disentangle the role of initial conditions from other drivers of growth, we further incorporate changes in technology, changes in frictions over time and other changes in the aggregate economy (such as time-varying population growth) in a standard general equilibrium Hopenhayn model of plant dynamics with endogenous entry and exit. In this framework, we are interested in the quantitative importance of initial conditions: how much aggregate growth is driven by simply letting initial conditions play out while keeping all other primitives fixed?

In the estimation framework, we show how the model is identified and how initial conditions are separately identified from other drivers of growth using standard plant- or firm-level micro data. The micro data is key to measure initial conditions, factor accumulation as well as productivity across plants over time. To this end, the data may show any level of initial misallocation and factor accumulation and is not required to be at a steady state at any point in time. We then apply the framework to study Indonesia's growth since 1975, using 40 years of panel data on Indonesian manufacturing plants. Besides the availability of exceptionally

[1] For Japan, see Chen, İmrohoroğlu, and İmrohoroğlu (2006). For European countries, see Alvarez-Cuadrado and Pintea (2009). For East Asian growth miracles, Young (1995) has argued that it reflects mostly factor accumulation, while recent work assigns a greater importance to TFP growth (Hsieh 2002; Lu 2012).

long-run and comprehensive micro data, Indonesia, the fourth most populous country in the world, shares many development patterns common to growing dynamic economies: a doubling of the population, a four-fold increase in the number of manufacturing plants, a 25-fold increase in manufacturing output, large sectoral reallocations and sizable changes in the plant size distribution over time.

To shed light on the drivers of growth in the Indonesian data, we start with some minimal, model-consistent structure. Heterogeneous plants produce output according to a common production function making input choices whose choice process we can remain entirely agnostic about. In this case, the following identity explicitly links heterogeneity across the plant distribution to aggregate output:[2]

$$
\begin{aligned}
\%\,\Delta\ \text{GDP} &\equiv \%\,\Delta\ \text{factor inputs}\ + \%\,\Delta\ \text{aggregate TFP} \\
&\overbrace{\%\Delta \begin{matrix}\text{aggregate}\\ \text{technology}\end{matrix} + \%\Delta\Big[\underbrace{\begin{matrix}\text{avg plant}\\ \text{productivity}\end{matrix}}_{\text{selection}} + \underbrace{Cov\begin{pmatrix}\text{plant} & \text{share of}\\ \text{productivity} & \text{inputs}\end{pmatrix}}_{\text{reallocation}}\Big]} \qquad (1)
\end{aligned}
$$

The first line captures the classic growth accounting identity that distinguishes factor inputs from aggregate TFP. Aggregate output can grow due to increases in factor inputs or due to increases in aggregate TFP. Distinguishing factor inputs from aggregate TFP is key to disentangle the accumulation and distribution of inputs from TFP improvements - a key step to isolate the role of initial conditions. This step requires disentangling plant-level TFP from observed inputs by estimating a plant-level production function. The second line further decomposes aggregate TFP, which in the presence of plant heterogeneity not only depends on aggregate technology but also on the selection of plants and the allocation of resources across heterogeneous plants. That is, aggregate TFP can be improved by reallocating resources from low productive to high productive plants. To separately identify changes in aggregate technology from changes in plant-level productivity, we exploit that aggregate technology improves the productivity of all plants in the economy, while selection only captures changes in the relative productivity of entering versus exiting plants. We can thus identify changes in aggregate technology from observed changes in productivity for incumbent plants after using the model to control for selection of surviving.

Quantifying each term of the accounting identity provides a summary measure of how aggregate technology and the entire joint distribution of plant-level productivity and resources changed over time and how these changes affected aggregate manufacturing output. We find five key results. First, most manufacturing

[2] The exact accounting identity and the underlying assumptions are given in Section 2.3. We derive the accounting identity starting from the fact that economic growth is simply the product of aggregating up the heterogeneous growth rates in value added output across all plants in the economy, a standard assumption on the separability of productivity in plant production functions and the decomposition based on Olley and Pakes (1996).

output growth is mechanically explained by the direct effect of input growth (68.5%). This input growth was importantly driven by an economy-wide reallocation of resources towards manufacturing. For example, the manufacturing employment share increased 2.5 fold over the 40 year period we study. Second, dispersion in the plant-size distribution grew over time as the right tail of the distribution filled up by incumbent plants growing slowly over time and the left tail of the distribution increased due to a large increase in entry of small plants. On net, high entry and lower exit rates led to a 4-fold increase in the total number of plants over the time period. Third, aggregate technology increased by roughly 50% over the time period 1975-2015. This accounts for less than 11.5% of the roughly 25-fold increase in total manufacturing output, but more than 80% of the increase in output per worker. Fourth, selection and reallocation effects explain the remainder, around 20% of the growth in total manufacturing output. Of these 20%, selection contributes slightly more than reallocation (56% vs. 44%). This means that (endogenous) selection of more productive plants contributed approximately as much to the increase in average TFP across plants as aggregate technology. This also highlights the importance of the exit margin to lead to a better selection of plants over time and is in line with empirical evidence by Asturias et al. (forthcoming) who show that periods of fast economic growth are disproportionately driven by the improved selection of plants. At last, for reallocation, we find that the positive growth effect is entirely captured by the mechanical increase in the number of plants. The contribution of the pure reallocation effect is actually negative in the data. That is, the allocation of resources as measured by the covariance of plant productivity and their share of resources actually worsened over time, mostly due to a large increase of productive entrants with few inputs. That is, entry likely undid some of the endogenous reallocation of resources across plants over time.

Initial conditions through the lens of the decomposition exercise play out via growth in factor inputs, changes in selection and changes in reallocation over time. However, the decomposition exercise does not capture indirect behavioral responses nor does it allow to separate different drivers of input growth, selection and reallocation effects. To fully quantify the role of initial conditions and other drivers of growth, we thus link the data to a quantitative model of endogenous plant dynamics that is still in line with the empirical decomposition results. We model manufacturing plants in a standard Hopenhayn model of plant dynamics with endogenous entry and exit to flexibly capture endogenous selection. To account for the (endogenous) increase in entry over time, we allow the mass and distribution of potential entrants to vary over time. We further allow for flexible changes in aggregate technology over time. Plants in the model face factor adjustment costs and financial constraints that can explain slow plant growth and the slow reallocation of resources over time. Plant growth also depends on the supply of inputs, which are partly endogenous in our model. The focus in the paper lies on human capital, because we observe it more consistently over time and because there is strong evidence that population growth and huge educational expansions have been key for Indonesia's development. While we take the aggregate increase in human capital (measured in efficiency units of labor) as given, we explicitly model the endogenous reallocation of labor from other parts of the economy to manufacturing by modeling a two-sector economy where labor can move across manufacturing and a rest-of-the-economy sector. This two-sector model is key to link output growth in manufacturing to economy-wide GDP and to explain dramatic changes in the manufacturing employment share over time.[3]

We estimate the model directly on micro data and only use the macro moments, including the growth decomposition results, for model validation. Similar to the decomposition exercise, the key of our estimation strategy is to be able to disentangle inputs, plant productivity and aggregate technology at the plant-level. Using production function estimation, observed prices and exit probabilities, we can then estimate many parts of the model without having to actually solve it. In the last step, estimation reduces to find the remaining parameters governing plant frictions that minimize the distance between model-implied and observed input choices across all producing plants in the economy. To separate initial conditions from changes in frictions over time, we estimate adjustment costs only based on data from the first two years and capture changes in frictions over time via time-varying wedges on labor demand and fixed costs that drive entry and exit. The estimated model replicates untargeted moments extremely well. Besides matching the long-run decomposition moments, we almost exactly match the entire time path of aggregate GDP, the time-varying marginal distributions of productivity and plant size as well as their time-varying joint distribution. We find large adjustment costs, especially large convex adjustment costs, that rationalize why few plants in the data grow large quickly and imply slow aggregate transitions that play out over many decades. These adjustment costs likely capture a variety of real-world frictions that prevent plants from growing quickly and they are entirely in line with systematic evidence of slower life-cycle growth of plants in developing countries (Hsieh and Klenow 2014). Together with evidence on differences in initial distributions - which we take directly from the data - adjustment costs explain the lack of large plants in developing countries (Hsieh and Olken 2014).

Using the estimated structural model, we quantify the role of initial conditions. We find that starting from the initial distribution of plants and fixing all primitives in the first year, the model predicts manufacturing output to roughly triple over time as plants grow, the initial plant distribution fills the right tail, more productive plants survive and resources are endogenously reallocated towards manufacturing. In our context, initial conditions explain 20% of the economy-wide output per capita gains between 1979 and 2015 that are due to changes in manufacturing. Transition dynamics are slow despite being dominated by reallocation

[3] We abstract from physical capital in our baseline results because of data limitations, but also estimate a model version with physical capital and very similar results in an extension.

dynamics instead of aggregate factor accumulation. We find that selection dynamics are relatively fast: about 90% of the the long-run average plant productivity is reached after the first decade, which is mostly driven by a convergence in productivity across plants. In contrast, entry and exit as well as reallocation dynamics are much slower. Most reallocation of resources takes more than 20 years and 40 years are far too little to come close to the stationary mass of plants. Importantly, initial conditions do not become meaningfully less important over time. The reason is that while initial conditions push the economy towards a new stationary distribution, other changes in the economy continuously shift this stationary distribution. Transition dynamics do not become less important over time if they are slower than changes in primitives that are moving the stationary distribution. We find strong evidence for this in the Indonesian data, driven mostly by sizable changes in aggregate technology and the distribution and mass of potential entrants. Specifically, we predict a similar long-run doubling of manufacturing output when starting from the observed plant distribution in 2015 and fixing all primitives in 2015.

The rest of the paper is structured as follows. Below, we discuss how our results relate to the literature. In section 2 we present the main empirical evidence. Section 3 builds on this empirical evidence to develop a model and discusses identification, estimation and validation. In Section 4, we use the structural model to formally decompose aggregate growth into its main drivers. The last two sections discuss robustness, extensions and policy implications.

Literature

We contribute to at least four different literatures. First and foremost, we contribute to the growth and firm dynamics literature in the tradition of H. A. Hopenhayn (1992). This literature has mostly focused on studying stationary equilibria (e.g. Restuccia and Bento 2015) or balanced growth paths (e.g. Midrigan and Xu 2014), abstracting from economic growth or transition dynamics. This omission is important, because we find that the data is far away from a steady state and transition dynamics are sizable compared to technology, the biggest driver of growth highlighted in the literature. This makes it difficult to think about policy and counterfactual development paths when abstracting from growth and transition dynamics. Important exceptions are Moll (2014), Song, Storesletten, and Zilibotti (2011), Poschke (2018), Buera and Shin (2013), and Buera and Shin (2017). Compared to these contributions, we do not only provide a much closer mapping between data and model, but we are also the first to study transition dynamics, firm heterogeneity and aggregate technology growth in combination. Our key result that transition dynamics do not become less important over time could only be shown in a model with these three building blocs.[4]

[4]More specifically, Buera and Shin (2013) are the first to study transition dynamics in a general equilibrium model with firm heterogeneity, but do not feature aggregate technology. Song, Storesletten, and Zilibotti (2011) captures aggregate technology,

Second, we contribute more generally to the quantitative macroeconomics literature by showing how to estimate the model directly on the observed transition path in the data, identify the model entirely on plant-level output and input choices and only use macro moments for validation. This allows for a much closer mapping between the model and observed growth paths in the data. Since a large part of the estimation can be done without actually having to solve the model, we gain computational tractability that we exploit; for example, by allowing for a non-parametric productivity process at the plant level and a more flexible parametric form for the plant adjustment costs than usually considered in the literature (e.g. Cooper, Haltiwanger, and Willis 2015). We also show how this estimation procedure holds for a variety of different equilibrium concepts related to long-run expectations: e.g. myopic, perfect foresight and constrained rational expectations equilibria. Given that we find little evidence for the anticipation of crises and our data includes both the Asian Financial crisis in 1998 and the Great Recession, we find myopic and constrained rational expectations equilibria a better representation of the data. This means our setup is very close to a standard business cycle setup with aggregate uncertainty (e.g. Clementi and Palazzo 2016). For our baseline results, we use a form of myopic long-run expectations that is in line with a form of constrained rational expectations and with stationarity in observed prices and aggregate technology in our context.[5]

Third, we contribute to the misallocation literature. Our paper follows a few recent papers that consider wedges in the tradition of Hsieh and Klenow (2009) in a dynamic setup David and Venkateswaran (2019). While we focus on adjustment costs as the only dynamic friction, we are, to the best of our knowledge, the first to consider wedges in a nonstationary setup. In this setup, we propose a general and tractable method that allows to identify changes in frictions over time from changes in aggregate labor market wedges over time. Conceptually, this allows us to distinguish between growth that is driven by continuously relaxing market frictions over time (e.g. Hsieh and Klenow 2009) and growth driven by the undoing of misallocation stemming from market frictions that were already relaxed (e.g. Buera and Shin 2013; Song, Storesletten, and Zilibotti 2011). We find the latter to be far more important in the Indonesian context.

At last, our paper relates to a literature on the decomposition of productivity and output. Our paper directly builds on Olley and Pakes (1996) for decomposing TFP into a selection and reallocation effect (also see: Melitz and Polanec 2015). These statistical decomposition exercises have been criticized for not being able to show causality, not capturing indirect effects and giving misleading interpretations of aggregate

initial misallocation, frictions and changes in these frictions over time, but only features two types of representative firms and no entry and exit. In contrast, our setup allows us to study selection and reallocation across the entire distribution of plants.

[5] To the best of our knowledge, this is the first paper to consider stochastic transition dynamics (for macro development, see e.g.: Buera and Shin 2013, 2017; Moll 2014). Given the potential importance of this contribution, we discuss constrained rational expectations in the tradition of Krusell and Smith (1998) at length. The key idea we exploit is that given the observed path of equilibria, we can directly estimate the forecasting rule in the data, avoiding a computationally costly iteration on parameters of the forecasting rule. This is very similar in idea to tractably solving perfect foresight equilibria taking observed prices as given as in Gopinath et al. (2017) or Caliendo, Dvorkin, and Parro (2019).

technology (e.g. Baqaee and Farhi 2020). In our paper, we show how the benefits of structural modeling and statistical decomposition exercises can be combined: using accounting identities to focus on what variation in the data is the most important and using the structural model to infer causality and quantify drivers of growth.

2 Empirical evidence

In this section, we introduce the Indonesian data and key facts about aggregate growth and changes in the distribution of plants over time. We then go through our main growth decomposition exercise as well as further empirical exercises that motivate the importance of initial conditions in driving growth.

2.1 Data

Our primary data comes from the plant-level Annual Manufacturing Survey (Survei Tahunan Perusahaam Industri Pengolahan), collected by Indonesia's Central Bureau of Statistics (Badan Pusat Statistik). 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 industry, employment, production and other plant characteristics from 1975 to 2015, a period of 41 years. It covers between roughly 6,500 to more than 25,000 plants per year. Using information based on a random five percent sample of all manufacturing establishments from the Indonesian Economics Census in 2006 reported in Hsieh and Olken (2014), 99% of Indonesian manufacturing plants have less than 20 workers, which our dataset entirely misses. On the other hand, most of the plants we miss are very small so that - based on the GGDC 10-sector database for Indonesia, which captures time-consistent aggregate sectoral employment and output series over the time period 1960-2012 (Timmer, de Vries, and De Vries 2015) - our manufacturing data captures about 30% of total manufacturing employment as well as about 30% of total manufacturing value-added output.[6] Given the data limitation for small plants, we can only look at 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 fraction of output in the overall economy, and we discuss how results would generalize in case we could observe more plants or more sectors.

[6] We 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 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 Economics Census, coverage is more complete after Economics Census years. We come back to this point in the next section.

For the empirical parts below, we draw on a yearly measure of the number of workers at a plant, which we take as the sum of total reported paid and unpaid workers. We further use the plant's yearly total wage bill (production + non-production wage bill), reported value-added output and reported per worker wages for production workers and non-production workers. Value-added output is not reported in 1976-1978, which forces us for both the growth decomposition exercise and the structural model to restrict ourselves to the time period 1979-2015. We also draw on a measure of a plant's capital stock based on Cali-Presidente (2021). This capital stock series draws primarily on self-reported capital stocks by plants, but drops observations that do not pass a battery of consistency checks. Missings are then filled up by drawing on the perpetual inventory method (PIM). Unfortunately, this capital series only starts in 1990. All variables denoted in Indonesian Rupiah are deflated to real values using the aggregate CPI.

2.2 Growth and changes in the plant size distribution over the long run

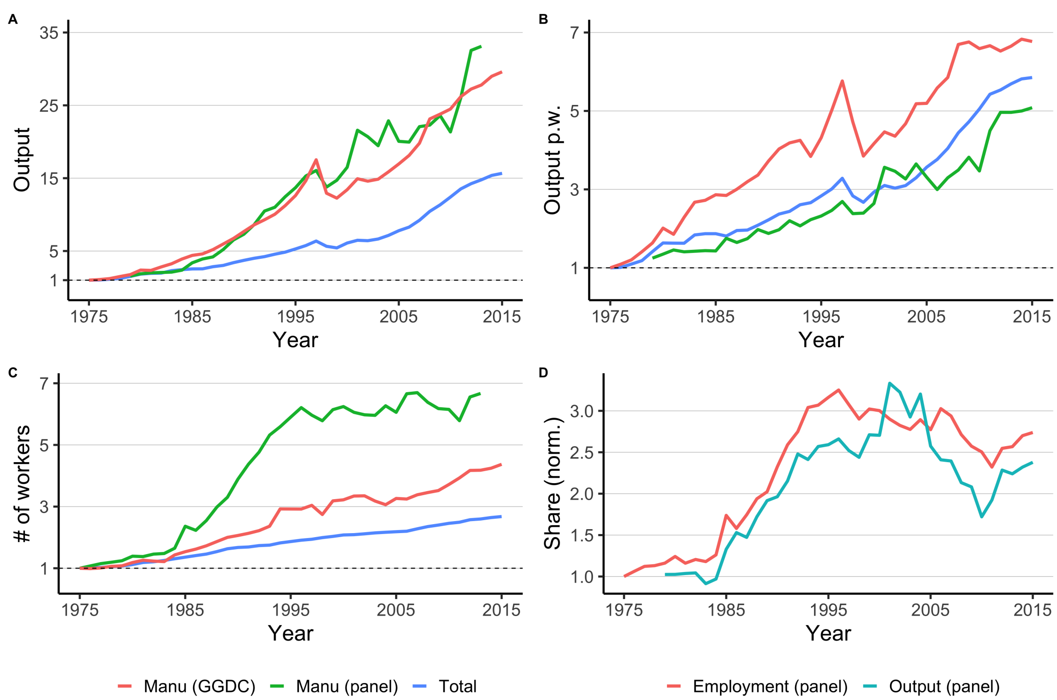

Figure 1: Evolution of aggregate and sectoral employment and output. Economy-wide data as well as one of the manufacturing series is based on the GGDC 10-sector database (1975-2012). Panel data refers to the manufacturing plants in the Indonesian manufacturing plant census (1975-2015). All series across all panels are normalized by their respective value in the first year to highlight increases over time. Panel A plots the increase in aggregate GDP and total manufacturing value-added output. Panel B gives the same as per worker versions. Panel C reports the evolution of the number of workers. Panel D gives the employment and output shares taking as totals the aggregates from the GGDC 10-sector database and the aggregates for manufacturing from the plant census.

Indonesia shares many development patterns common to growing dynamic economies. Figure 1 reports the evolution of employment and output at the aggregate and manufacturing-sector level over the period 1975-2015. Based on the GGDC 10-sector database, aggregate GDP in Indonesia grew by a factor of more than fourteen between 1975 and 2012, the working population increased by a factor of more than 2.5 and hence GDP per worker increased more than 5-fold. Manufacturing contributed importantly to this aggregate growth: manufacturing output grew 25-fold and employment reallocated majorly to manufacturing as shown by a 2.5-fold increase in the manufacturing employment share.

To understand what drove the 25-fold increase in output in manufacturing, we document large changes in the plant size distribution over the same time period. We start out by plotting the average plant size over time in Figure 2 Panel A, where size is simply measured by the number of workers. The average number of workers in medium and large manufacturing plants in Indonesia roughly doubled between 1975 to 2015.[7] Figure 2 Panel B shows that the number of plants in the panel roughly increased four-fold over time. One can clearly see that the large drops in the average size of plants coincide with the inclusion of new and smaller plants around 1985, 1996 and 2006. Thus, the data gives a census of manufacturing plants with more than 20 employees around the time of the Economic Census, but fails to be a complete census in the intermediate time periods as not all new plants are included over time. However, despite year-to-year fluctuations being partly driven by specific attributes of the survey sampling, the long-run picture of increasing average plant size and entering plants that are smaller clearly stands out.

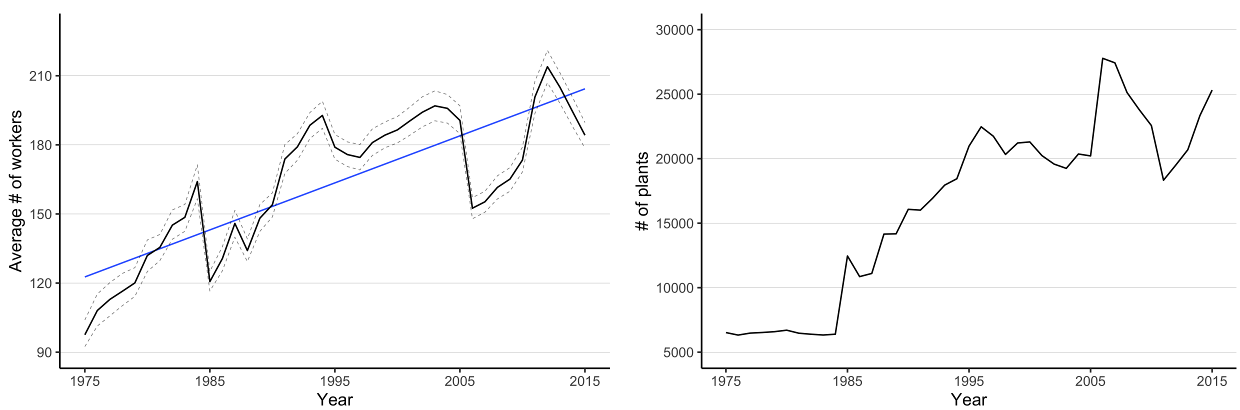

Figure 2: Panel A (left): Evolution of average plant size measured in number of workers (paid + unpaid) for medium- and large Indonesian manufacturing plants. Solid black line gives point-estimate, grey dotted lines give 95% bootstrapped confidence intervals and solid blue line gives best linear fit. Panel B (right): Evolution of number of medium- and large Indonesian manufacturing plants in the panel. Data: Based on Statistik Industri, the Indonesian manufacturing plant census capturing plants with 20 or more workers. Restricting sample to plants with at least 20 workers (enforce cutoff across survey years) and less than 15k workers. Total of 88,282 unique plants, ranging from 6,797 in 1975 to 26,311 in 2015.

[7] The series further shows substantial fluctuations with larger drops in 1985, 1995 and 2006. The reason for this is that these years are years of the Economic Census in which the plant panel was filled up with plants that were previously missed. These plants are mostly small plants, either because they are young (and thus are newly recorded in the census), because they are small (close to the cutoff of 20 workers) or both.

Next, we consider changes in the entire plant size distribution in Figure 3. A number of features stand out. First, the plant size distribution has a heavy right tail across time, clearly visible under the log-scale in Panel A. Secondly, the plant size distribution moves to the right. This is most clearly visible by focusing on the evolution of right tail moments as shown in Panel B. The plant at the 75th percentile almost doubles its size of 67 workers in 1975 to 125 workers in 2015. At the 90th percentile, plants more than double their size from around 175 in 1975 to around 400 in 2015. This increasingly heavier right tail is also what drives the increase in the average plant size. At last, the dispersion of the plant size distribution increases over time as the 25th percentile stays roughly constant around 26 workers and the higher percentiles move further to the right.

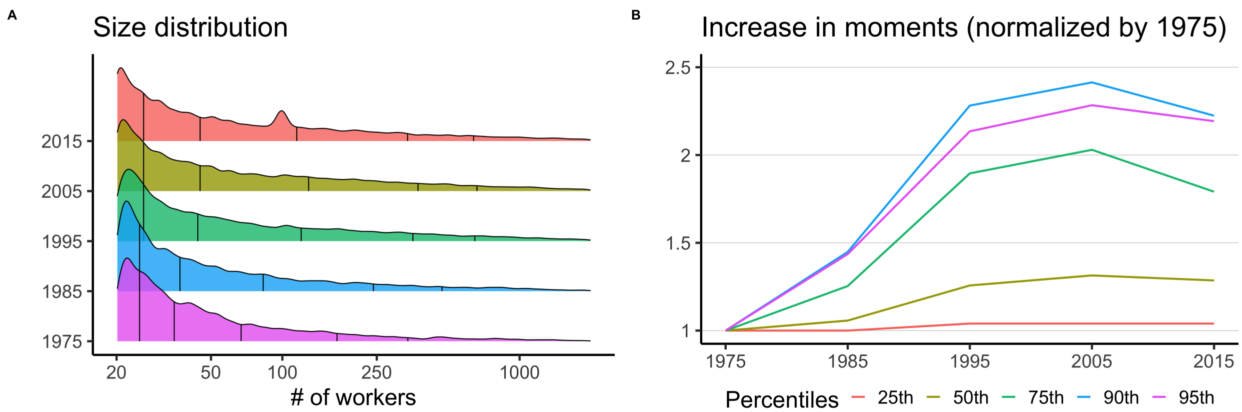

Figure 3: Evolution of distributions of total number of workers (paid + unpaid) for medium- and large Indonesian manufacturing plants. Panel A (left) gives the evolution of the entire size distribution, vertical lines give 25th, 50th, 75th, 90th and 95th percentiles. Panel B (right) plots separately the evolution of the same percentiles, normalizing each by their value in 1975. Data is based on Statistik Industri, the Indonesian manufacturing plant census capturing plants with 20 or more workers. Restricting sample to plants with at least 20 workers (enforce cutoff across survey years) and less than 15k workers as well as to data from 1975, 1985, 1995, 2005 and 2015.

The increasing importance of large plants over time can also be seen by focusing on their total employment share. Figure 4 shows yearly employment shares in large plants by computing for each year the total employment in plants of a certain size over total employment across all plants in the panel. The employment share in plants above 500 workers increases steadily from roughly 40% to more than 60% of total employment. Similarly, for plants with at least 1,000 workers, the employment share almost doubles from around 23% to around 44%. Large plants become much more prevalent and important over time and are largely responsible for the increasing average size of manufacturing plants in Indonesia. Most of this increase is concentrated in the period up until the mid-1990s.

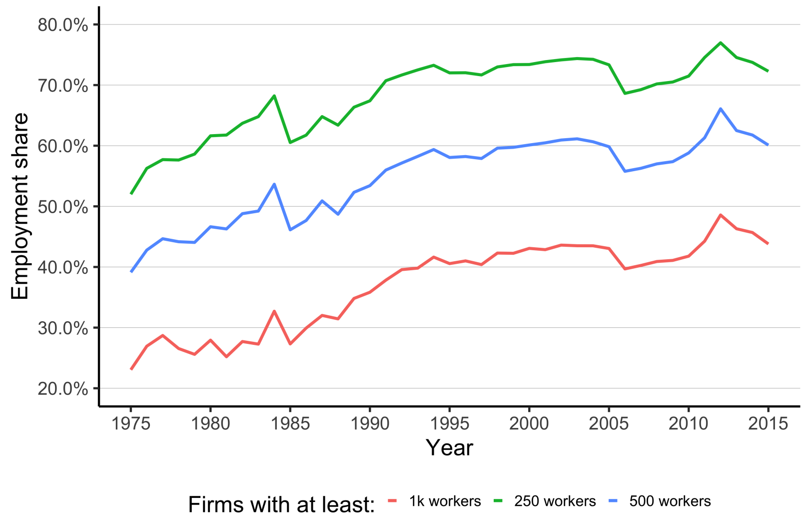

Figure 4: Evolution of employment share in large Indonesian manufacturing plants. Based on Statistik Industri, the Indonesian manufacturing plant census capturing plants with 20 or more workers. Restricting sample to plants with at least 20 workers (enforce cutoff across survey years) and less than 15k workers. Total of 88,282 plants, ranging from 6,797 in 1975 to 26,311 in 2015. Details on the variable construction in the text.

2.3 Decomposing aggregate growth

To better understand Indonesia's aggregate growth over the last 40 years, we start out with a formal decomposition. Throughout this section, we focus solely on manufacturing using our plant-level data. This statistical exercise does not show causality, but it does reveal which features in the data are crucial to explain aggregate manufacturing growth. Note that by definition, total output Y_t is the sum of value added output y_{it} across all manufacturing plants:

$$Y_t \equiv \sum_{i \in \Omega} y_{it}$$

Next, we assume a standard plant production function of the following form:

$$y_{it} = z_t * s_{it} * f(x_{it}) \tag{2}$$

where z_t captures aggregate technology, the central driver of economic growth in 'Neoclassical' growth models, whether it is endogenous or treated as being exogenous. s_{it} in turn captures plants' idiosyncratic productivity. At last, $f(x_{it})$ is an increasing function in inputs x_{it} such as labor and capital. Given this general setup, we can write total output as:

$$lnY_t \equiv ln \sum_i f(x_{it}) + lnTFP_t \equiv ln \sum_i f(x_{it}) + lnz_t + ln \left[\bar{s}_t + N_t cov \left(s_{it}, \frac{f(x_{it})}{\sum_i f(x_{it})} \right) \right] \tag{3}$$

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 can then be further decomposed into aggregate technology and a combination of average productivity and a covariance term that captures whether resources in the economy are allocated towards the most productive plants. This last decomposition of a productivity term into an average and a covariance term is based on Olley and Pakes (1996). We differ from the decomposition literature here in that we further distinguish plant productivity from aggregate technology and that we focus on aggregate output instead of aggregate TFP. Growth in aggregate output is then given by:

$$\Delta\, ln\, Y_t \equiv \underbrace{\Delta\, ln \sum_i f(x_{it})}_{\text{input growth}} + \underbrace{\Delta\, ln\, z_t}_{\text{aggr techn.}} + \underbrace{\Delta\, ln \left[\bar{s}_t + N_t cov\left(s_{it}, \frac{f(x_{it})}{\sum_i f(x_{it})}\right)\right]}_{\text{selection + reallocation effect}} \tag{4}$$

To quantify each term's contribution to growth in aggregate output, we need to separately identify each component of the plant-specific production function. First and in line with the subsequent model, we assume that plants produce with the same standard Cobb-Douglas production function $f(x_{it}) = k^{\alpha} h^{\theta}$ using capital and labor as production inputs. We measure labor in efficiency units by dividing the firm's wage bill by a measure of the real wage.[8] As discussed in the Data section, we do not consistently observe capital over the entire time series, since the capital series only starts in 1990 and capital stocks are notoriously difficult to measure. In the baseline results below, we thus abstract from capital. However, in the Appendix, we report decomposition results where we (1) focus only on the time period from 1990 to 2015 for which we observe capital and (2) where we impute capital using information on capital from after 1990 and enforcing model-consistent optimal capital choices. Results are broadly similar and we discuss their differences below.

We estimate the output elasticity θ allowing for labor to be fully dynamically chosen. Our most preferred estimator does not rely on the static choice of intermediate inputs as traditionally considered in the production function estimation literature (e.g. Ackerberg, Caves, and Frazer 2015; Olley and Pakes 1996) and we discuss it in more detail in the structural estimation part and in the Appendix. The idea is that in general settings where plants observe today's productivity but face dynamic frictions when making input choices (e.g. adjustment costs or financing frictions), the output elasticity in combination with first-order conditions makes sharp extensive margin predictions on whether plants should increase or decrease their inputs compared to the previous period. The estimator we propose is agnostic about the actual dynamic frictions and estimates θ only based on correctly classifying whether plants increase or decrease their number of workers. Based on this estimator, we find that $\hat{\theta} \approx 0.58$, considerably larger than the average observed labor share in our data (≈ 0.45) and closer to labor shares estimated in developed countries. Given $f(x_{it})$ and observable

[8] We discuss the identification and estimation of the wage series in the model part.

output y_{it}, plant total factor productivity (TFP) is identified by: $TFP_{it} \equiv s_{it}z_t = \frac{y_{it}}{f(x_{it})}$. We estimate plant TFP using reported plant-specific value-added revenue and efficiency units of labor.

We are interested in identifying changes in aggregate technology z_t. For this, we assume that z_t and s_{it} are uncorrelated and we normalize s_{it} such that any drift in the productivity process is part of z_t. There are two main issues in separating z_t and s_{it}. The first issue relates to the selection of producing plants and changes in their composition. For example, if less productive plants are more likely to exit, we will observe a positive drift in average TFP across time even in the absence of any change in aggregate technology z_t.[9] While this drift in average TFP is going to be important for accounting for aggregate output, it is driven by a selection effect and not by aggregate technology z_t. A second and less studied issue is when the distribution of active plants is not at their ergodic productivity distribution (assuming it exists). For example, suppose there is no aggregate technology growth, no entry and exit and we observe a sample of plants whose idiosyncratic productivity evolves according to a bounded AR(1) process with persistence parameter $|\rho| < 1$. Suppose all plants start at the lower bound of productivity, then the subsequent time path of average productivity will show a positive drift as plants move closer to the ergodic distribution.

We deal with both issues by using simply the median within-plant change in TFP:

$$\widehat{\Delta ln(z_t)} \equiv \text{Median}_{i\in S}\left(\Delta ln(TFP_{it})\right) = \Delta ln(z_t) + \text{Moment}_{i\in S}(\Delta ln(s_{it})) \tag{5}$$

where S is the subset of surviving plants.[10] This estimator allows for arbitrary entry and allows exit to be correlated with inputs as well as contemporaneous productivity: that is, smaller and less productive plants are allowed to be more likely to exit (as is true in the data). However, the estimator rules out that exit is correlated with future productivity realizations.[11] This setup is consistent with the subsequent structural model that we use. Also, we find that current changes in inputs are only very weakly correlated with future productivity changes for surviving firms, indicating little selection on future productivity realizations. A benefit of having a model is that we can also directly compute the model-based bias term $\text{Moment}_{i\in S}(\Delta ln(s_{it}))$ to verify that it is zero. We indeed find a zero yearly bias when using the median, but find large biases when using the average.[12]

[9]We can write this estimator as: $\Delta\overline{log(TFP_t)} = \Delta log(z_t) + \frac{1}{N_t}\sum_{i\in S_t} log(s_{it}) - \frac{1}{N_{t-1}}\sum_{i\in S_{t-1}} log(s_{it-1})$, which is only an unbiased estimate in case the average idiosyncratic productivity across active plants in t equals that of active plants in $t-1$. This assumption is violated if for example plant exit is correlated with a plant's idiosyncratic productivity or where entry changes over time and new entrants have different productivity. Both of which hold in the data, invalidating the use of this estimator.

[10]In the Appendix, we also build and discuss alternative estimators using different moments than the median and different subsets of active plants.

[11]Technically, the estimator is unbiased in the case where $\text{Median}_{i\in S}(\Delta ln(s_{it})) = 0$. This will be violated in models where plants first draw a signal about future productivity before making their current exit choice.

[12]Specifically, the model-based bias term depends on first separately identifying z_t and s_{it}, then discretizing idiosyncratic

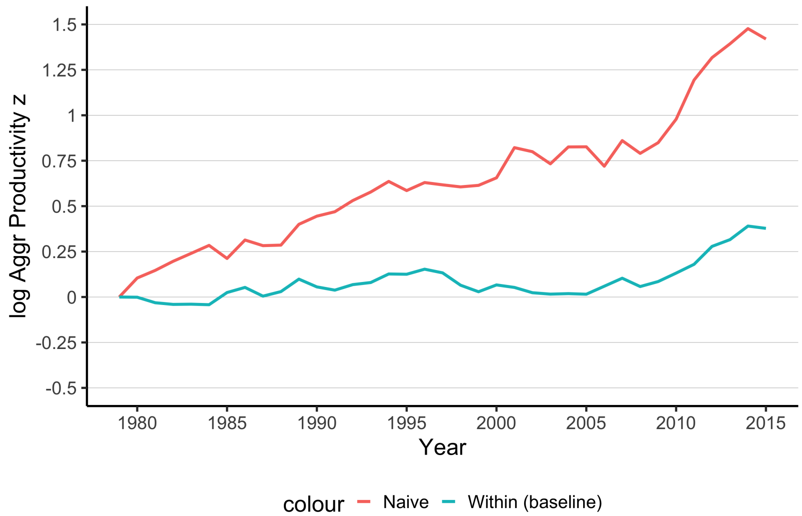

Figure 5: Evolution of aggregate technology (in logs) for two different estimators. Initial values are normalized to unity. Naive gives the estimator based on average TFP across plants (not controlling for selection). Within (baseline) gives the baseline estimator that additionally controls for selection of plants. Plant-level data based on Statistik Industri. 33,904 unique plants and 296,202 plant-year observations.

In Figure 5 we plot our baseline estimator of aggregate technology z_t and compare it to the "naive" estimator of simply taking the average TFP across all plants over time. Given that the level of aggregate technology is unidentified, without loss of generality we normalize the initial value to unity. We find that average plant TFP increased strongly over time, but that aggregate technology only accounts for a fraction of this growth. That is, ignoring plant selection gives a very misleading picture of aggregate technology growth. Still, we find a sizable growth in aggregate technology, making plants with the same idiosyncratic productivity s 46% more productive in 2015 compared to 1979.

Given identification of all production function parts, we can finally fully decompose manufacturing output growth for each year in Indonesia from 1979-2015 using Equation (4). Since we are interested in the drivers of growth over the entire time period, we compare the year 1979 to 2015. We find that the contribution of aggregate technology growth to aggregate output growth in manufacturing is at most 11.9%. In contrast, labor accumulation makes up around 68.5% of the total and selection plus reallocation accounts for the remaining 19.5%. Note that labor accumulation here includes not only an increase in the total efficiency units of labor, but it also captures an increase in the number of plants as well as the distribution of labor across plants. Due to decreasing returns to scale in production, it is beneficial to have labor distributed more evenly across plants.

productivity, non-parametrically estimating its process and finally drawing s_{it} from this process for surviving plants with idiosyncratic productivity s_{it-1} to reconstruct the bias. The median is powerful here especially because of the discretization: most plants do not move bins even if the underlying productivity distribution is not at the ergodic distribution so that the median changes are zero, while average changes may show a strong drift/bias.

Further decomposing selection and reallocation, we find that selection is slightly more important than reallocation (56% vs. 44%). We show this by taking into account that the terms enter non-linearly into the formula. To quantify their relative contribution from t to $t+1$, we fix respectively one of the terms at their value at t and only change the other, then add both contributions and compute their relative shares. Looking within reallocation, we can further distinguish the pure reallocation effect (the covariance term) from the mechanical effect that stems from changes in the number of plants. We find that the pure reallocation effect actually contributes negatively (-74%), while the increase in the number of plants has a strong positive effect (174%). That is, reallocation actually worsened over time. The reason for the negative reallocation effect is that the covariance between plant productivity and plant output shares declined over time because of the massive entry of small but productive plants and slow plant growth leading to a slow reallocation of resources. Another likely reason that contributed to the negative reallocation effect is that labor supply increased dramatically and estimated wages remained fairly constant, ameliorating the pressure for less productive plants to let workers go and reallocate these to more productive plants.

The main takeaways from this exercise, and which any structural model need to replicate, are (1) that the direct effect of aggregate technology on aggregate growth is fairly small, (2) that there is a substantial amount of input growth as plants' average size as well as the number of plants grows, (3) that there is sizable positive selection on productivity as less productive plants exit and entering and surviving plants become more productive, and finally (4) that reallocation of resources in the economy is slow and on net negative because of the large increase in the number of entrants. A large benefit of the decomposition exercise is that these results hold irrespective of dynamic frictions in the economy and plant behavior; at this point, we have only made an assumption about the production function and on the identification of aggregate technology.

In the Appendix, we provide further details on this exercise. First, we show in more detail how the relative contribution of each of these factors changed over time for each year in the data. Relatedly, we further decompose parts by studying the relative contributions of entry and exit. Secondly, we study the robustness of the decomposition results by considering alternative estimates of aggregate technology and by considering changes in the contributions from changing the output elasticity θ. We also discuss at length how results change when considering capital. While the quantitative numbers do change, the main insights stay the same.

2.4 The importance of initial conditions and slow transitions

The previous framework gives a systematic view on the direct drivers of economic growth, but it does not capture indirect drivers nor does it give causality. For example, why was reallocation of factors of production from low productive to high productive plants so slow over the 40 years of study? Or what drove the huge observed increase in factor inputs? While at least part of it is indirectly driven by aggregate technology, both could have been driven by a variety of other policy changes that removed distortions over time and steadily increased plant entry and input growth. In this subsection, we provide reduced-form evidence that the seeds of plant growth were already present at the beginning of our sample in 1975 and that future plant growth is highly predictive. We take this as evidence for the importance of initial conditions and slow transitions and against a large role for future policy changes - a key result that also holds in our quantitative model.

Specifically, we consider the following exercise. Take as the starting point the initial size distribution of plants Φ_t in 1975 as measured by the number of workers and discretized in X different size bins. Each bin captures the fraction of plants with this number of workers. Additionally, we can follow individual plants and compute the probability of moving from one bin to the other between periods t and $t+1$, which we summarize in the transition matrix $P_{t,t+1}$ of dimension X^2.[13]

Both factor accumulation and reallocation of resources could be driven by productive plants needing time to grow large plants. In practice, various constraints may prevent productive plants from directly choosing their right size including input adjustment costs and financial frictions. To study the potential importance of slow transitions, we use the transition matrix $P_{t,t+1}$, which we fix, to simply iterate on the initial distribution. We view the resulting predicted changes in the plant size distribution over time through the lens of two metrics: the average plant size and the share of workers working in plants with more than 500 workers.

Results for this exercise using $X = 10$ and observed transitions only from the first two years in the data are reported in Figure 6. Starting from the initial plant size distribution in 1975 and only feeding in information of transitions between 1975 and 1976, we get a long way in explaining changes in the entire plant size distribution over the subsequent 40 years. More than 70% of average plant size increases (A) and more than 90% of the increase in the employment share of plants with more than 500 workers (B) are explained by simply iterating on the initial distribution. Together, these two results also help explain what is likely driving these transition dynamics. In 1975, the distribution lacks large plants and based on the transitions observed between 1975 and 1976, we can rightly predict that the share of large plants will increase slowly over time until it converges somewhere close to 60% of total employment; the stationary share based on

[13] At this point, we abstract from entry and exit behavior, that is we restrict ourselves to plants who employ workers in both period t and $t+1$. Alternative exercises with entry and exit are reported in the Appendix.

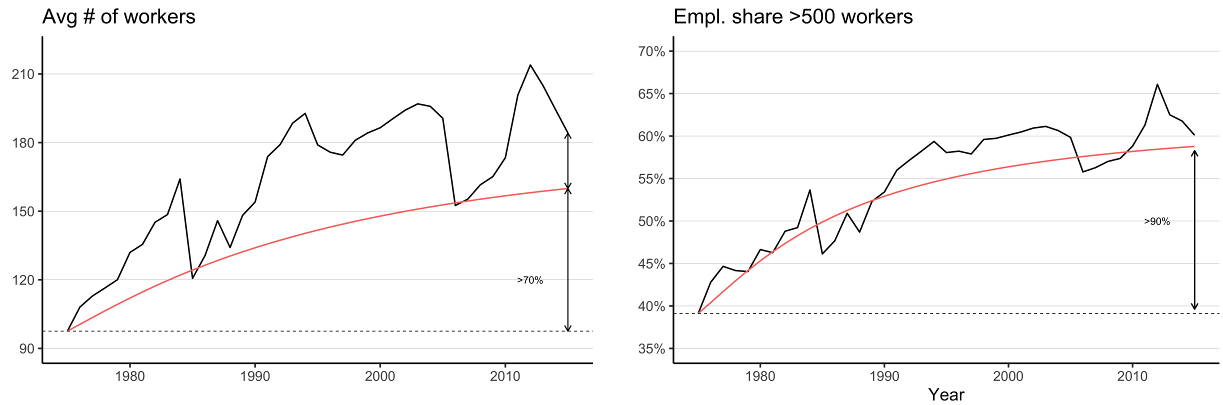

Figure 6: Reduced-form transition dynamics implied by initial plant size distribution. Black line gives the data and red line gives predicted changes over time. Details are in the text.

the transition matrix $P_{1975,1976}$. Since the initial employment share is far below 60%, the average plant size increases. In the Appendix, we provide a battery of robustness exercises to these results, including varying initial distributions and transition matrices, averaging transition matrices over multiple years and varying assumptions on entry and exit behavior. Furthermore, we show very similar results for real value added output and the labor wage bill.

We report further evidence in Figure 7. In line with the idea of the initial distribution of plants being away from their stationary distribution, Panel A documents that most plants grow their labor shares over time.[14] Panel B plots the observed distribution of labor shares against our estimated output elasticity for labor. In a static and frictionless setup, the optimal plant size is given when the labor share equals the output elasticity on labor. Again, this plot shows that there is considerable scope for plant growth as most plants are below this optimum. In the cross-section, we do not see a systematic movement towards this optimum over time, because (1) the composition of plants changes due to the entry of small plants and selection of productive plants, and (2) the increase in aggregate technology that reduces the labor share for everyone. These two forces prove to be key drivers of transition dynamics in our quantitative model.

Taken together, this section has provided evidence for aggregate growth being importantly driven by initial conditions and slow transitions. These slow transitions are driven by some form of frictions that prevent plants from growing too quickly and the combination of entry and aggregate technology growth ensures that even over longer periods of time there remains a strong potential for future plant growth as summarized in the distribution of plants. We now turn to a structural model that can account for these patterns and flexibly allows for an important role of initial conditions and the plant size distribution, and where slow plant

[14] We take the median within-plant growth in labor shares by plant age, pooled over time and weighted by previous plant size. We exclude plants for which labor shares more than doubled or halved between any two periods to ensure that our results are not driven by outliers. This result is robust to different moments and outlier corrections.

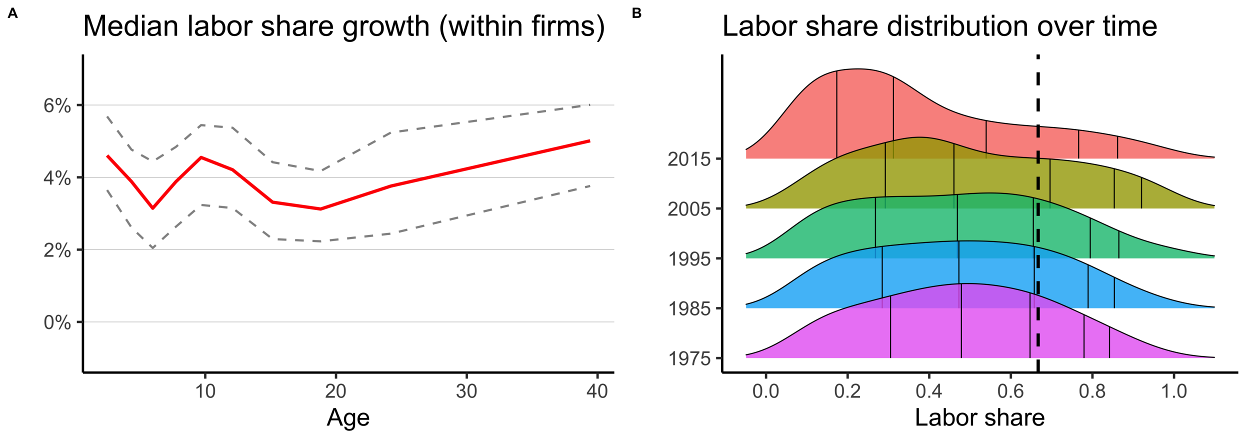

Figure 7: Within-plant and cross-sectional evidence on labor share evolution. Panel A: Evolution of within-plant labor share growth over plant age. Estimates give the weighted median across plants by plant age, weighted by the previous number of workers. Furthermore, estimates are pooled across all years (N = 394,397) and by 10 equal-sized age groups. Plot gives non-parametric bootstrapped 95 percent confidence bands. Further details in the text. Panel B: Evolution of cross-sectional labor share distributions over time. Dotted black line gives estimated output elasticity.

growth is driven by adjustment costs. This gives a more direct interpretation of the decomposition exercise and endogenizes the transition matrix $P_{1975,1976}$ as well as changes in this transition matrix over time.

3 Structural model of aggregate growth and plant dynamics

We consider a standard model of heterogeneous plant dynamics in the tradition of H. A. Hopenhayn (1992) featuring endogenous exit and entry. Plant size evolves endogenously due to the combination of time-varying exogenous aggregate technology and idiosyncratic productivity, endogenous exit and slow endogenous hiring of labor that is affected by adjustment costs and financial frictions. We then embed this model of plant heterogeneity into a two-sector general equilibrium model to study economy-wide growth processes and endogenous reallocation of resources across sectors.

3.1 Model Setup

Sectors

The model economy is set in discrete time indexed by $t = 1, 2, \ldots$. There are two sectors of production: Manufacturing (M) and a rest-of-the-economy sector (R). Both sectors produce the same homogeneous good. The manufacturing sector features heterogeneous plants, while the rest-of-the-economy produces the good as a representative plant with a decreasing returns to scale (DRS) production function:

$$y_t^R = A_t \left(h_t^R\right)^{\theta_R} \quad \text{with} \quad \theta_R \in (0,1) \tag{6}$$

where A_t is time-varying TFP, h_t^R gives labor employed in the rest-of-the-economy as measured in efficiency units and θ_R gives the output elasticity for sector R. 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: $\pi_t^R(A_t, w_t, \tau_t) = y_t^R(A_t) - w_t h_t^R$, but faces labor demand wedges τ_t^R that distort first-order conditions such that optimal labor demand is given by:

$$h_t^{R*} = \left(\frac{\theta_R A_t}{(1+\tau_t^R) w_t}\right)^{\frac{1}{1-\theta_R}} \tag{7}$$

For simplicity, we assume that both A_t and τ_t^R follow deterministic exogenously given processes that flexibly capture changes in technology adoption and labor frictions over time. Labor markets in both sectors are fully competitive (up to wedges) and there is a single wage w_t that clears both labor markets.

Manufacturing plants

In manufacturing, the model features a mass of heterogeneous, risk-neutral plants that produce y_{it} at time t according to the production function introduced in the previous section:

$$y_{it} = z_t s_{it} \left(h_{it}^M\right)^{\theta} \tag{8}$$

As discussed previously, we abstract from capital here due to data restrictions. Efficiency units of labor h are the only factor of production, whose output elasticity is captured by θ. Production depends on exogenous aggregate technology in manufacturing z_t and exogeneous idiosyncratic manufacturing productivity s_{it}, which are assumed to be independent. Both processes are potentially highly persistent. Idiosyncratic manufacturing productivity is allowed to follow any general bounded Markov process of order one, nesting nonlinear and non-Gaussian processes. Assumptions on aggregate technology in manufacturing depend on the equilibrium concept we use and can vary from assuming a fully flexible but deterministic exogenous process to a persistent log-linear Gaussian autoregressive process.

Plants choose inputs based on their previous input use due to the presence of adjustment costs, which they pay if they want to change their input use. Labor is assumed to be hired on a spot market. We model adjustment costs in a flexible way following the literature Cooper, Haltiwanger, and Willis (2015). Formally, labor adjustment costs AC are given by:

$$AC_h(h_{i,t-1}, h_{i,t}) = \begin{cases} F^+ + c_0^+(h_{i,t} - h_{i,t-1}) + \frac{c_1^+}{2}\left(\frac{h_{i,t}-h_{i,t-1}}{h_{i,t-1}}\right)^2 h_{i,t-1} & \text{if } h_t > h_{t-1} \\ 0 & \text{if } h_t = h_{t-1} \\ F^- + c_0^-(h_{i,t-1} - h_{i,t}) + \frac{c_1^-}{2}\left(\frac{h_{i,t}-h_{i,t-1}}{h_{i,t-1}}\right)^2 h_{i,t-1} & \text{if } h_t < h_{t-1} \end{cases} \tag{9}$$

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, which are also allowed to be asymmetric. 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. These convex costs could be organizational in nature. For example, there might be limits to training new workers or the human resources department might have limited capacity and increasing this capacity has large short-run costs by having to reallocate labor, retraining workers, etc. Alternatively, there could be additional costs from collective action (in the case of firing) or it might simply be hard to locally find many suitable job candidates at once. These convex adjustment costs will be key to explain slow growth of plants over time.

On top of adjustment costs, plants face financing constraints that further constrain their labor demand choices. We choose a tractable cash-flow based financing constraint that arises from a working capital constraint. Specifically, we assume that plants need to hire (and pay) their workers before production realizes, forcing them to borrow their wage bill. There are financial intermediaries who commit payment to plant owners and commit to workers that their wages will be paid. However, financial intermediaries enforce a simple financing constraint that ensures plant owners have no interest in running away with the borrowed wage bill. This financing constraint becomes:

$$y_{it}(s_{it}, h_{it}, z_t) - w_t h_{it} \geq 0 \tag{10}$$

$$\kappa\left(y_{it}(s_{it}, h_{it}, z_t) - w_t h_{it}\right) \geq w_t h_{it} \tag{11}$$

where κ measures the tightness of the financing constraint. For simplicity, we assume the time plants need to borrow goes to zero such that plants face zero interest costs on the financing. In the Appendix, I provide more details on this borrowing constraint, related literature and evidence for such cash-based borrowing constraints in the data.

The timing of manufacturing production is summarized in Figure 8. Denote by Ω_t the aggregate state of the economy, which is important for plant decisions and will be made more precise later on. Incumbent plants

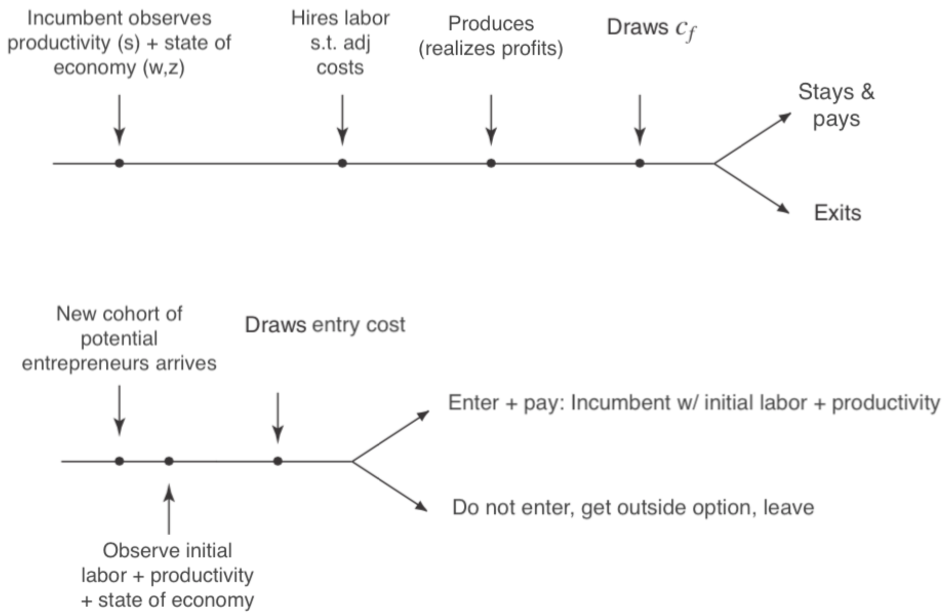

Figure 8: Timing in time t for manufacturing firms.

choose current inputs to produce and generate per period profits:

$$\pi(s_{it}, h_{i,t-1}, h_{i,t}; z_t, w_t) = y_{it}(s_{it}, z_t, h_{i,t}) - w_t h_{i,t} - AC_h(h_{i,t-1}, h_{i,t}) \tag{12}$$

After production takes place, incumbent plants incur a fixed cost of production c_f, drawn from a distribution G based on which they decide whether they want to continue producing or permanently exit (as in Clementi and Palazzo 2016). Note that we define profits before paying the fixed costs. The value of exiting is given by an outside option b that captures any opportunity costs of production and the costs of closing down the plant (as in H. Hopenhayn and Rogerson 1993). For now, we leave these costs of closing down unspecified and simply denote them by exit costs$(h_{i,t})$. The decision problem of a plant with cost draw c_f is then:

$$\max\left\{\frac{1}{R}\mathbb{E}[V^M(s_{i,t+1}, h_{i,t}; \Omega_{t+1})|s_{i,t}, h_{i,t}; \Omega_t] - c_f, b - \text{exit costs}(h_{i,t})\right\} \tag{13}$$

where V^M gives the value of an incumbent plant. The survival decision then reduces to the probability $\lambda(s_{i,t}, h_{i,t}; \Omega_t)$ that the operating cost draw c_f is lower than the future expected value of an incumbent manufacturing plant: $\mathbb{P}(x \geq c_F) = G(x)$ where $x \equiv \frac{1}{R}\mathbb{E}[V^M(s_{i,t+1}, h_{i,t}; \Omega_{t+1})|s_{i,t}, h_{i,t}; \Omega_t] - b + \text{exit costs}(h_{i,t})$.

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

$$V^M(s_{i,t}, h_{i,t-1}; \Omega_t) = \max_{h_{i,t} \in [\underline{h}, \bar{h}]} \Bigg\{ \pi(s_{it}, h_{i,t-1}, h_{i,t}; z_t, w_t) + \lambda(s_{i,t}, h_{i,t}; \Omega_t) \left\{ -\mathbb{E}_c[c_F|\text{stay}] + \frac{1}{R}\tilde{\mathbb{E}}_{s,w,z}\left[V^M(s_{i,t+1}, h_{i,t}, \Omega_{t+1})\right] \right\} \Bigg\} \tag{14}$$

where $\tilde{\mathbb{E}}_{s,w,z}$ denotes expectations over future prices, aggregate and idiosyncratic manufacturing productivity. The tilde on the expectations operator, $\tilde{\mathbb{E}}$, makes clear that plants, depending on the equilibrium concept chosen, may deviate from perfect rational expectations here. Plants discount profits at rate $1/R$, where R is the exogenously given international interest rate that is assumed to be constant over time. The presence of adjustment costs, financial constraints and fixed costs makes this a dynamic problem where plants take into account that contemporaneous changes in their inputs will influence adjustment and financing costs and the ability to pay fixed costs in the future.

At last, we consider endogenous plant entry. As visualized in Figure 8, each period there is a cohort of potential entrants (PE) of measure $|PE|_t$ who each draw a random entry cost c_E from a distribution P, which they need to pay in case they want to start producing. Their outside option is also given by b. Potential entrants differ in their idiosyncratic productivity (which they know at time t) and their initial labor $h_{i,t}$. The distribution of potential entrants is given by $PE_t(h_t, s_t)$, which is potentially time-varying due to exogenous reasons such as population growth and specifics of how plants are measured in our data, a point we further discuss in the Appendix. The initial level of inputs is also exogenous and captures an initial firm size which the potential entrant can reach without facing adjustment costs and one can think of this as labor supplied directly by the entrepreneur and her family and friends. A potential entrepreneur i then chooses whether to enter or not according to:

$$V_{PE}(s_{i,t}, h_{i,t}; \Omega_t) = \max \left\{ V^M(s_{i,t}, h_{i,t}; \Omega_t) - c_E, b \right\} \tag{15}$$

Similar to exit, this gives the following mapping $\mathbb{P}\big(V^M(s_{i,t}, h_{i,t}; \Omega_t) - b \geq c_E\big) = P\big(V^M(s_{i,t}, h_{i,t}; \Omega_t) - b\big)$. Importantly, with this specification the initial mass and distribution of entrants is endogenous, but entrants only start making input choices the period after they entered.[15] 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. With slight abuse of notation, denote by $\mu_t(\Omega_t)$ both the entire distribution and total mass of entrants at time t. Similarly, define by $m(h_t, s_t)$ the mass of producing plants for each state (h_t, s_t) in period t and by M_t the entire distribution and total mass

[15] We also tried out a model version where entrants start with inputs in $t-1$ and already make input decisions at the time of entry t. For technical reasons related to the ability to invert observed entry for inputs at $t-1$, we chose this version.

of producing plants at time t.

Households

The economy is populated by a mass of hand-to-mouth households j who consume their labor income and who inelastically provide labor supply each period. The aggregate resource constraint is given by households inelastically consuming all final products $C_t = Y_t = \sum_{i \in M} y_{it} + Y_t^R$.[16] Furthermore, households may have idiosyncratic efficiency units of labor $h_{j,t}$ which aggregate to total labor supply: H_t. Aggregate labor supply may vary due to exogenous population growth, exogenous changes in the composition of workers and exogenous changes in average skills and schooling, which we all treat as deterministic processes, so that we denote the exogenous sequence of aggregate labor supply by $\{H_t\}_{t=0}^{\infty}$. Importantly, households allocate their labor supply across sectors based on maximizing labor income. Thus, the labor supply elasticity that manufacturing plants face changes over time due to exogenous changes in the total labor supply and changes in labor demand in the rest of the economy that are in part driven endogenously by the prevailing wage rate.

Equilibrium

The focus in this paper is on a (growth) path of per-period *Recursive Competitive Equilibria.* Denote by $Z_t \equiv \{A_t, \tau_t^R, H_t, PE_t\}_t^{\infty}$ the sequence of exogenous deterministic processes in the economy starting at t. The aggregate state space then captures the state of aggregate technology, Z_t and the endogenous mass of producing firms M_t: $\Omega_t \equiv \{z_t, Z_t, M_t\}$. We define a *Recursive Competitive Equilibrium* by Ω_t and an endogenous wage w_t such that:

1. The rest-of-the-economy sector statically chooses optimal labor demand maximizing profits taking productivity A_t and the wage rate w_t as given and facing labor demand wedges τ_t^R.
2. Manufacturing plants choose optimal labor demand and optimal output by maximizing profits, taking as given previous labor demand, s_t, z_t, w_t and Ω_t and forming expectations $\tilde{\mathbb{E}}_{s,w,z}$ over future s_{it}, z_t and w_t.
3. Potential entrants optimally choose entry based on knowing the future value of incumbents.
4. Households inelastically supply total labor H_t and optimally allocate labor across sectors to maximize labor income.
5. The aggregate wage w_t adjusts to ensure that the labor market clears: $H_t = h_t^R(w_t, A_t, \tau_t^R) + \sum_{i \in M_t} h(s_{it}, h_{i,t-1}; z_t, w_t, \Omega_t)$

[16] Note that we implicitly assume that entry, exit and adjustment costs are not actual resources but only disutility.

6. The mass of active plants in t and previous aggregate state Ω_{t-1} is equal to surviving plants from $t-1$ plus endogenous new entrants:

$$\forall(s_t,h_t):\quad m(s_t,h_t;\Omega_t)=\sum_{s_{t-1},h_{t-1}}\Big(\mathbb{1}_{h^*=h_t}\mathbb{P}[s_t|s_{t-1}]\lambda(s_{t-1},h_{t-1};\Omega_{t-1})\times m(s_{t-1},h_{t-1};\Omega_{t-1})\Big)+\mu(s_t,h_t) \tag{16}$$

7. The goods market clears each period: $C_t=Y_t=\sum_{i\in M}y_{it}+Y_t^R$

This definition of a *Recursive Competitive Equilibrium* nests different equilibrium concepts based on the assumption on expectations: rational expectations for the case of $\tilde{\mathbb{E}}_{s,w,z}=\mathbb{E}_{s,w,z}$, constrained rational expectations equilibria in the tradition of Krusell and Smith (1998) in case $\tilde{\mathbb{E}}_{s,w,z}$ features forecasts for the wage w, perfect foresight in case of certainty over aggregates (w,z), and different forms of myopic equilibria.

3.2 Solving the model and identifying changes in frictions

The model features a flexible role for initial conditions as well as continuous changes in the entire economy over time (time-varying aggregate technology in manufacturing and the rest-of-the-economy, time-varying labor supply, plant entry and time-varying labor market frictions in the rest-of-the-economy). The endogenous path of labor market equilibria in the model depends on an entire distribution of plants that changes endogenously over time. To allow for a tractable analysis of this economy - both along the observed equilibrium path and for counterfactual paths - we make a simplifying choice on expectations and provide new technical tools regarding the estimation and computation of the model.

The main simplifying step we take is to consider a *Recursive Competitive Equilibria* with a form of constrained rational expectations. Following the literature, we deviate from the idea that plants track the entire distribution of plants and can perfectly forecast how the current distribution of plants as well as all other future aggregate shocks will affect future prices. We show how to tractably solve either for a path of perfect foresight equilibria (assuming there is no aggregate uncertainty) or for constrained rational expectations equilibria in the tradition of Krusell and Smith (1998). The implementation of Krusell and Smith (1998) in a non-stationarity setting is a novel and important technical contribution. However, in the end, we see the choice of expectations as an empirical question and we find little evidence for forward looking expectations over future aggregate shocks. For the baseline results, we thus further simplify from expectations over aggregate risk and model plants as being only forward-looking about idiosyncratic risk (productivity, entry and fixed cost shocks), but have myopic expectations over aggregate changes in the economy. This means that at time t and given the wage w_t and aggregate technology z_t, plants expect this wage and technology to stay constant over time. This greatly simplifies the model computation by allowing for a sequential computation of equilibria and provides a great fit to the data. We relegate the estimation of perfect foresight and constrained rational expectations equilibria to an extension where we additionally show that there exists a form of constrained rational expectations that is consistent with our baseline expectations.

On the computation and estimation of the model, we gain tractability by taking a revealed price approach commonly used in Empirical Industrial Organization, but applied here to a path of time-varying equilibria. The "standard Macro" way of solving our model would be to solve for the entire time path of equilibria for each set of parameters and find the set of parameters that brings the model closest to the data based on a set of well-chosen moments (e.g. via simulated method of moments). The main problem with this approach is that solving for each path of equilibria is computationally costly, preventing flexibility on the amount of model parameters that are estimated. Following recent ideas in the estimation of macroeconomic models (e.g. Gopinath et al. (2017) & Caliendo, Dvorkin, and Parro (2019)), we exploit the fact that in the data we observe one path of equilibria. The idea is to invert the model computation by not trying to solve the model for a path of prices that clears markets, but instead using the observed price path in the data and finding the set of parameters that are consistent with the observed path of equilibria. The key reason why this allows for a more flexible but still simpler estimation of model parameters is that for many sets of model parameters, the model never actually has to be solved. Specifically, our approach allows us to non-parametrically estimate idiosyncratic productivity, directly estimate realizations of aggregate productivity, the time-varying distribution of potential entrants as well as flexibly estimate parameters for adjustment costs.

At last, a key input into the computation and estimation of the model are time-varying labor market wedges. These wedges fulfill a dual role. First, we introduce time-varying labor market wedges that unexpectedly shift plants' labor demand to ensure that observed prices that we enforce in the estimation actually clear labor markets in our model economy. Dealing with market clearing is a central, though little discussed issue in recent estimation methods for equilibrium models across different areas in Economics (cite!!). The approach in the recent literature that enforces observed prices - which is most often not made explicit - is to either not actually guarantee market clearing or allow model supply and demand to deviate from observed supply and demand (cite!!). Here, we introduce wedges to make this connection explicit. These wedges capture a measure of our ignorance: if the model would perfectly capture the data, these wedges would be zero for each period. Importantly, we enforce the same wedges in counterfactuals, ensuring that counterfactuals are comparable to the baseline estimates.

For the second role of wedges, we reinterpret them to disentangle the role of initial conditions versus changes in policy over time. Similar in spirit to the reduced-form exercise in Section 2, we first estimate the model parameters related to frictions present in the first two years in the data and then interpret changes in wedges over time as changes in frictions. The idea is simple: If frictions estimated from the initial years weaken over time, we would expect to see more plant growth over time than predicted by our model and this would be captured by positive wedges. One can think of this as a wedge accounting exercise that extends the static setup in Hsieh and Klenow (2009) to a dynamic and non-stationary setting. In principle, we can identify these time-varying wedges at the state space level as in Hsieh and Klenow (2009). However, given that the plant distribution evolves endogeneously over time, we cannot guarantee that the model-implied distribution of plants over the state space stays the same as in the data even if we enforce state-specific wedges.[17] While we still compute time-varying state-specific wedges as an additional model validation step, we instead opt for a much simpler aggregate wedge of the form:

$$h_{i,t} = (1 + \tau_t \omega_{s,t}) h^*(s_{i,t}, k_{i,t-1}, h_{i,t-1}; z_t, w_t) \tag{17}$$

$$\omega_{s,t} \equiv \frac{\bar{h}_{s,t}^{\text{Data}}}{\bar{h}_{s,t}^{\text{Model}}} \tag{18}$$

where τ_t is a uniform wedge across all manufacturing plants in a given period that ensures that model-implied labor demand aggregated across the endogenous distribution of incumbent plants equals aggregate labor supply in the data. $\omega_{s,t}$ is a productivity-year-specific weight that ensures that the distribution does not diverge over time. Specifically, it could be that for a given year, aggregate demand and supply equalize despite the model not capturing well the distribution across productivity and size. In this case, the weights ensure that for every year where there is a non-zero wedge, the distribution of plants is brought back closer to the true distribution. This issue of a deteriorating distribution affects any estimation of heterogeneous agent models with non-stationary dynamics and we provide a new tool to deal with it.

3.3 Identification and Estimation

In the following, we go through all parameter identification and estimation steps. We distinguish direct estimation, which allows estimation without having to solve a model first, and indirect estimation, which requires to solve (part of) the model.

[17] One reason is that endogenous exit will not exactly be equal to exit in the data. Also, we update plants based on our estimated productivity process, which does not ensure that each plant updates its productivity as observed in the data.

Parameterization of standard parameters

We start out with setting $\beta = 0.95$ for all households in our economy as well as the rest of the world. Based on standard arguments, this implies that the international interest rate is given by $R = \frac{1}{\beta}$. All the remaining parameters of the model are estimated.

Direct estimation of productivity and wage processes

Next, we proceed with the direct estimation part. We start with the manufacturing sector. In line with the analyses in Section 2, we use θ to directly identify TFP: $TFP_{it} \equiv z_t s_{it} = \frac{y_{it}}{h_{it}^{\theta}}$. The labor bill is observed in the data, which we divide by an estimate of the real wage to obtain efficiency units of labor: $h_{it} = \frac{\text{wagebill}_{it}}{w_t}$. We identify the real wage in the data by drawing on reported real per worker wages in the manufacturing survey. To capture true changes in the real wage, we need to make sure that these wages capture remuneration for the same job and the same worker skills. Ideally, we want to capture changes in the real wage for a worker whose efficiency units of labor remained constant. In the absence of worker-level data, we focus on average changes in reported real per worker wages within plants who have seen little changes in their workforce. For this, we focus on plants whose total number of workers changed by less than 5% between t and $t+1$ (for the smallest plant with 20 workers in the dataset, this allows a maximal change of one worker). Furthermore, we focus only on reported production worker wages in contrast to other workers and we exclude plants whose per worker wages increased or decreased by more than 50% between any two time periods. At last, we use the weighted average of wage growth across plants, weighting by the number of production workers within a plant. After normalizing the initial wage to unity, Figure 9 plots the estimated real wage series in the data.

Next, we estimate θ. As discussed in Section 2, we propose a novel estimator that allows labor to be fully dynamically chosen. While we explain technical details of the estimator in the Appendix, the key idea behind the estimator comes from looking at the policy functions implied by a model with adjustment costs as shown in Panel A of Figure 10. Each separate line gives the policy function for a specific idiosyncratic productivity realization in period t. Each dotted line gives the static unconstrained optimal choice of labor, which is independent of previous labor. Note that the dotted line in combination with the 45 degree line (which captures whether plant inputs grow or decline) separates the policy space into four regions. Importantly, optimal model-implied policies only happen to be in two of the four regions, which is a general result across many dynamic frictions. In case a plant is above their optimal static unconstrained labor demand (as judged by current productivity and past labor demand), a plant always wants to weakly decrease their labor demand, while in the opposite case, a plant wants to weakly increase labor demand. The estimator exploits

Figure 9: Evolution of the real wage in Indonesian manufacturing using within-job changes in observed real wages. Estimation is based on reported wages of production workers from plant-level data based on Statistik Industri.

this geometric result by finding the θ that maximizes correctly classifying plants into these two regions and minimizing the number of plants that are misclassified as being in one of the other two regions. Based on this estimator, we find that $\hat{\theta} = 0.577$ with bootstrapped 95% confidence bands ranging from 0.52 to 0.626.

We separately identify aggregate technology changes as in Section 2 and plotted in Figure 5. To see that this identification strategy works in the model here, remember that the model features only selection on current productivity, not future realizations of productivity. Most importantly, note that plants in the model may be far away from their optimal size, but that only shows up in the labor component and not in TFP and will thus not bias the estimation of aggregate technology. Having identified TFP and aggregate technology, we back out idiosyncratic productivity s_{it}, discretize it and estimate its transition matrix $\mathbb{P}(s'|s)$ non-parametrically by using observed frequencies in the data, assuming solely that s_{it} follows a bounded first-order Markov process.[18] Specifically, we choose 30 grid points for idiosyncratic productivity, which we select based on quantiles of the productivity distribution, oversampling highly productive plants to correctly capture the right tail of the plant size distribution.

Direct estimation of the state space, entry and exit

Given productivity and wages and normalizing their initial values to unity, we can directly identify any plant in the economy on the state space of our model. To identify the state space on a manageable grid for model

[18]See Ruiz-García (2019) for the quantitative importance of using a flexible productivity process rather than a standard log-normal process in structural models of plant dynamics. We also enforce minimal regularity conditions to ensure reasonable estimates of the transition matrix. These are discussed in the Appendix.

computation, we also discretize efficiency units of labor h_{t-1} by choosing 1000 grid points that we choose based on equal spaced quantiles, ensuring that the entire labor distribution is well represented. This allows a very close mapping between model and data that we exploit both for estimation and for the model validation. As a starting point, we directly identify the initial distribution of plants in 1979 over $(s_{i,t}, h_{i,t-1}; w_t, z_t)$.[19]

We can also separately look at entering and exiting plants to obtain potential entrant distributions and inform the model estimation of entry and fixed cost parameters that govern endogenous entry and exit. For exit, we can in principle directly and non-parametrically estimate the survival rate $\lambda(s, h, z, w)$ from the data. However, given that exit rates are only noisily observed in the data, we choose a standard parametric approach for the fixed cost distribution $G(x)$. Specifically, we assume that fixed costs are type-1 extreme value distributed with parameters μ_X and β_X. This has the benefits of having a closed-form expression for the CDF as well as the conditional expectation of the fixed costs $\mathbb{E}_c[c_F|\text{stay}]$. The closed-form expression allows us to analytically solve for μ_X and β_X by inverting two observed exit probabilities in combination with the model-implied expected continuation value V^M. That is, we obtain μ_X and β_X as a direct function of other model parameters that govern V^M and two chosen exit probabilities over the state space in the data. We choose exit probabilities at the 50th and 90th percentile of the plant size distribution in 1979 and their corresponding average productivity. We estimate these probabilities using the average exit rate around these percentiles. In practice, this means that we find the corresponding μ_X and β_X for each set of parameters governing financial frictions and adjustment costs.

For entry, we directly identify time-varying entrant distributions $E(s_t, h_t; w_t, z_t)$. Potential entrant distributions are then given by $PE_t(s_t, h_t) = E_t(s_t, h_t)/\mathbb{P}_E(s_t, h_t; w_t, z_t)$, where $\mathbb{P}_E(.)$ gives the entry probability for given parameters of the entry cost distribution. Since potential entrant distributions and the entry cost distribution are not separately identified, we impose that the entry cost distribution equals the fixed cost distribution. This is similar to normalizing the distribution of potential entrants and more general than normalizing the total number of potential entrants as often done in entry models.

Remaining direct estimation (labor supply + Rest of the Economy)

Total labor supply H_t is given exogenously and is simply 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 observed by aggregating up all observed h_{it} for each time period t. To obtain h_t^R, we use the fact that we observe the total number of

[19] We take 1979 as the initial year, because we observe labor of these plants in 1978, and 1979 is the first year for which we observe value added output, which we need to estimate idiosyncratic productivity. Strictly speaking, value-added output is also observed in 1975, but not between 1976-1978. We cannot use the 1975 value-added data, because we do not observe efficiency units of labor in 1974.

workers l_t^R in the Rest of the Economy. To map from the number of workers in R to the total efficiency units of labor in R, we directly draw on recent estimates of worker selection for Indonesia by Hicks et al. (2017). Specifically, we use their estimates of wage differences and worker selection across rural agriculture and urban non-agriculture as a benchmark for our two sectors. This leads us to estimate that average efficiency units of labor are roughly two times larger in M than in R in our economy and we use this to infer total labor supply in the rest of the economy: h_t^R.[20] For the rest-of-the-economy sector, 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. This gives $\theta_R \approx 0.226$. Wedges τ_t^R are backed 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: $A_t = \frac{y_t^R}{\left(h_t^R\right)^{\theta_R}}$.

Indirect estimation

The last part of the model estimation is based on indirect estimation. The remaining plant-side model parameters are the parameters governing financial frictions an labor adjustment costs $\Theta = \{\kappa, F^+, F^-, c_0^+, c_0^-, c_1^+, c_1^-\}$. Those cannot be directly inferred from the data and depend on a dynamic labor choice. This means that for estimation, optimal model-implied labor demand has to be solved for each combination of the parameters in Θ. We estimate the parameters at the plant-level using non-linear least squares (NLS) by minimizing the distance between model-implied labor demand and actually observed labor demand according to:

$$\hat{\Theta} = argmin\Big\{ \sum_i \Big(log(h_{1980}^{model}(\hat{s}_{i,1980}, h_{i,1979}; \hat{z}_{1980}, \hat{w}_{1980})) - log(h_{1980}^{data}(\hat{s}_{i,1980}, h_{i,1979})) \Big)^2 \Big\}$$

(where we might additionally weight observations by their labor demand to ensure that optimal parameters allow reasonable aggregation). We visualize this approach in Figure 9 below.

At last, we restrict the estimation to the first two periods of observed labor demand. As discussed before, this allows us to distinguish the role of transition dynamics from changes in frictions over time. Specifically, we fix the policy environment in 1979 and 1980 to estimate financial frictions and adjustment costs and introduce time-varying model wedges that ensure market clearing over time. We interpret these time-varying wedges as capturing a combination of IID noise and a potentially persistent "policy shock" component that is related

[20] 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.

to financial frictions and adjustment costs changing over time. To separate changes in frictions related to labor demand from changes in frictions related to entry and exit, we similarly allow the level of fixed costs μ_t to vary over time. We do so by finding the time-varying μ_t that ensures the right mass of plants in each period. We show all estimated wedges below.

Identification of the financial frictions and adjustment cost parameters is explained in Figure 10. Panel A gives the intuition of the identification, where a model-implied optimal policy is plotted based on some combination of parameters κ, F, c_o, c_1 (assuming symmetry for expositional clarity). Each line gives a different productivity realization. Intuitively, the inaction region given by the part of the policy function that is on the 45 degree line is governed by the fixed cost F and the hiring and firing cost c_0. In contrast, the curvature of the adjustment off the 45 degree line is captured by c_1. The higher these convex adjustment costs, the slower adjustments and the closer to linear these adjustments (the less curvature). Fixed costs F are theoretically identified from observed changes in the workforce for small plants, while variable costs c_0 are identified from larger plants for whom the fixed cost plays a much smaller role in the decision-making process. The financial frictions κ restrict behavior for plants that are much below or much above their optimal size, forcing them to grow much more slowly than they would like or shrink in size (hence the linear curves at the extremes).

Panel B in contrast highlights the actually estimated policy function by showing the fit against the data for a specific productivity shock realization. In the data, observed policies bunch around the inaction region in line with the presence of important adjustment costs. Furthermore, observed policies are a lot noisier than what is implied by the simple model with labor adjustment costs. What is crucial here is that the data clearly rules out strong curvature in labor adjustments. Most labor changes are close to the inaction region. In the model, this is only possible in the presence of high convex costs of adjustment.

3.4 Evaluating model fit

Given the estimation and computation of the model, we want to look at how well it fits the data. Our approach is from "micro to macro", whereby all estimation is based on micro moments and validation of the model is on non-targeted macro moments. In the Appendix, we have a longer discussion on validation and many more validation tests. Here, we focus on the main results. Figure 11 shows how our baseline model fits the time-varying path of the four main endogenous components of the formal growth decomposition exercise: output, inputs, selection and reallocation. The model closely tracks not only the long-run 24-fold increase in manufacturing output, but also yearly variation in output over time. This is only in part driven by feeding in the exact series of aggregate shocks, as becomes clear when looking at the decomposition of output. As

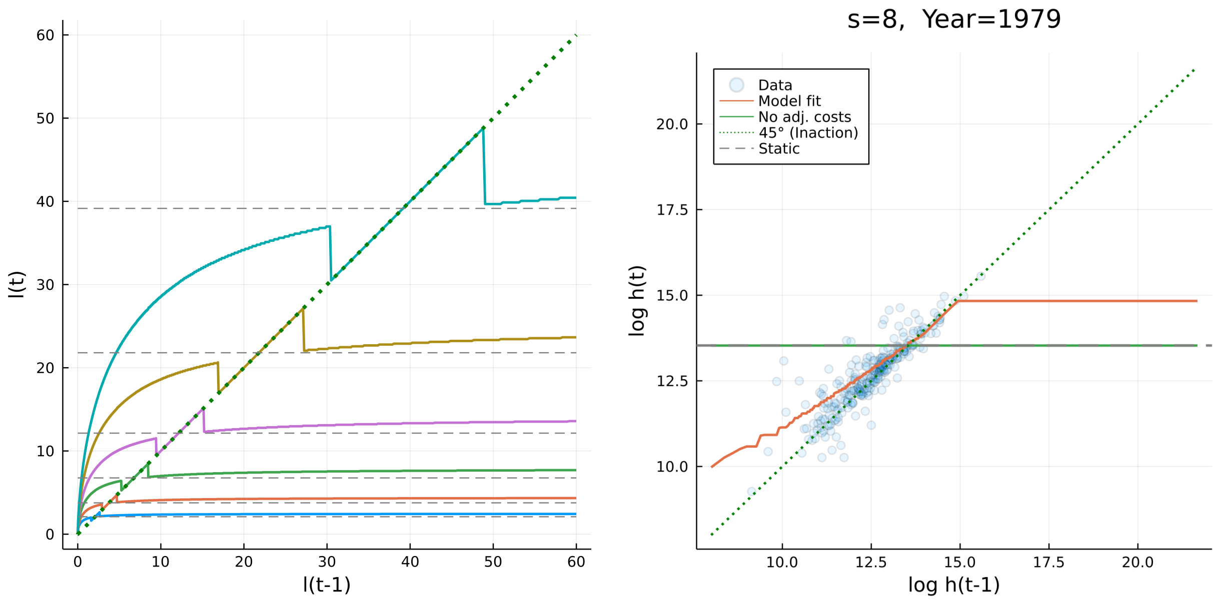

Figure 10: Identification and estimation of the adjustment cost parameters. Panel A gives a hypothetical optimal policy function based on some combination of parameters F, c_o, c_1 and no financial frictions. Each line gives a different productivity realization. The 45 degree line gives the inaction region in which plants do not change their plant size. Panel B gives the estimated policy function together with observed data for a specific productivity shock realization.

the main component of aggregate output growth, the model closely tracks the distribution of labor across the endogenous plant distribution over time. Here, total labor demand and the total number of plants is hit by construction, but the entire distribution of labor across plants is endogenous and left unrestricted. The model captures the right degree of slow labor accumulation across the entire plant size distribution. The model also tracks well the endogenous selection of plants over time. This is in part driven by allowing for a non-parametric productivity process that leads to a convergence of average productivity across plants, but in large part also driven by the endogenous selection of plants. Shocks in the series are due to entry shocks around census years, which the model gets right by construction, but the model also gets right the convergence and evolution of productivity after entry shocks by matching which plants are more likely to exit over time. At last, the model performs well on the most difficult part: the endogenous evolution of the joint distribution of productivity and size, as captured by the reallocation plot. Here, the model captures the long-run increase in the misallocation of resources, in part driven by entry and in part driven by slow reallocation dynamics. The model only really departs from the data in the initial years as the model predicts too much exit for small and productive and too little exit for large and unproductive plants.

Next, we can look at the evolution of wedges that are needed to ensure labor market clearing in each period as well as variation in the level of fixed costs μ_t that ensures the right mass of plants over time. Figure 12 shows that aggregate labor market wedges are on average positive and around 6%. That is, on average labor

Figure 11: Baseline model fit, comparing model results (blue) and data (orange). All graphs report results for manufacturing only and each panel represents one term in the main output decomposition formula. Top-left panel reports value-added output, with initial year normalized to 1. Top-right panel reports a summary measure of the labor distribution. Bottom-left reports average plant productivity. Bottom-right panel reports the covariance between plant labor and productivity.

demand by plants in the data is around 6% higher over time than what initial frictions and labor market conditions imply. This is not driven by a secular increase in wedges, but rather by comparatively low wedges in the first two years; if anything, we see a decrease in wedges over time, indicating that frictions slightly increase over time. The key driver of this pattern is that the model predicts aggregate technology growth to lead to slightly more plant growth than observed in the data. This could be driven by a worsening of plant frictions, or alternatively, simply by a slight failure of the model to correctly capture the elasticity of plant growth to aggregate productivity changes across the plant distribution. Furthermore, we see spikes in wedges over time, indicating temporary shocks in (reported) labor demand that is unrelated to aggregate technology changes or other measured changes in the economy over time. Moreover, the evolution of wedge weights over time shows that the model generally underpredicts dispersion in labor demand across the productivity distribution, predicting too much growth for low productive plants and too little growth for high productive plants. Importantly, there seem not to be any specific trends in these weights over time that would indicate important time-varying drivers of labor demand that the model misses. At last, the model captures well predicted exit over time. To see this, Figure 13 shows that there is little systematic variation in the level of the fixed cost distribution that is needed to exactly capture the level of exit over time.

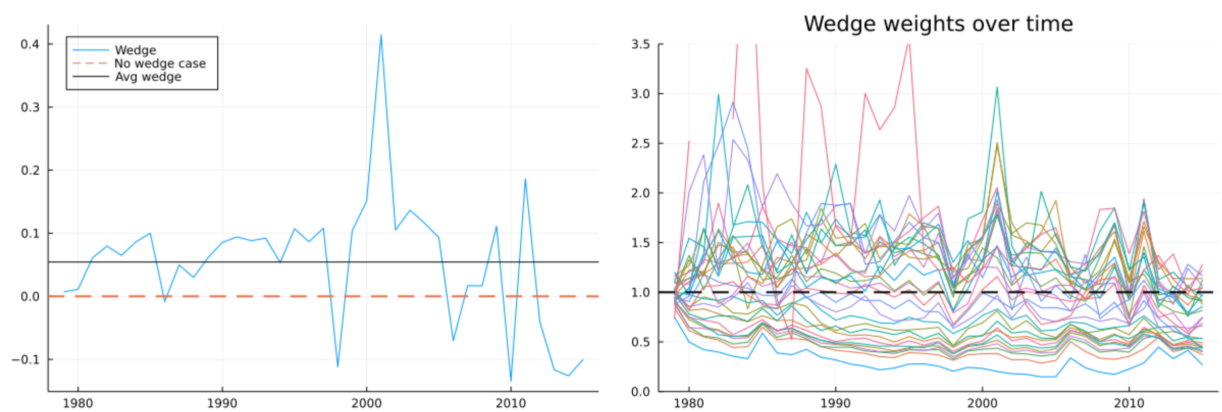

Figure 12: Estimated policy wedge and wedge weights by productivity state that ensure labor market clearing each period.

4 Quantifying the drivers of aggregate growth

In this section, we use the structural model to better understand the role of historical persistence and initial conditions in driving economic growth over time. We also use the model to speak to alternative drivers of economic growth such as aggregate technology and changes in frictions.

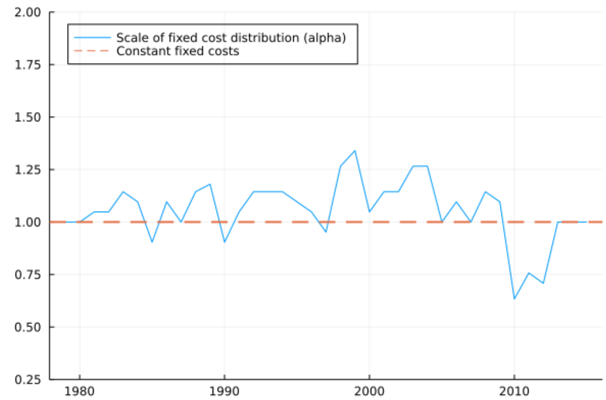

Figure 13: Estimated level of fixed costs that ensures the right total mass of plants over time.

Initial conditions and the role of historical persistence

Using the structural model, we find that initial conditions explain about 20% of observed growth over the period 1979-2015. To derive this number, we start from the 1979 economy: the initial distribution of plants and primitives from 1979 that are given by the potential entrant distribution, frictions, wedges, aggregate technology in manufacturing and the rest of the economy as well as aggregate labor supply.[21] We then consider a counterfactual equilibrium growth trajectory where the economy and distribution of plants evolves endogenously over time, but primitives stay constant over time. That is, we consider deterministic transition dynamics starting from primitives in 1979. We focus on aggregate output per worker as the main outcome of interest and measure of welfare, which allows to compare welfare across different counterfactuals with very different population growth. We make one important normalization such that results are interpretable. While the focus is on growth in aggregate output, manufacturing (as captured in our data) makes up less than 10% of output in Indonesia. This means that any change in manufacturing (e.g. different initial conditions or different technology growth) will have a relatively small effect on aggregate GDP. For this reason, we report results after taking out the GDP per worker increases that would have happened even in the absence of any changes in manufacturing.[22]

Figure 14 explains how initial conditions contribute 20% to observed growth. Total output in manufacturing

[21] We find that results are sensitive to variation in the initial distribution of potential entrants as this governs the long-run increase in the number of plants. For this reason, we take an average of the potential entrant distributions of the first few years until the next economic census year.

[22] Let $y_t^k \equiv \frac{Y_t^k}{L_t^k}$ be aggregate output per worker at time t in counterfactual k. We specify the contribution of counterfactual k as: $\text{Contribution}(k) \equiv \left(\left(y_{2015}^k - y_{1979}^k\right) - \left(y_{2015}^{k,\text{No manuf}} - y_{1979}^{k,\text{No manuf}}\right)\right) / \left(\left(y_{2015}^{\text{Baseline}} - y_{1979}^{\text{Baseline}}\right) - \left(y_{2015}^{\text{No manuf}} - y_{1979}^{\text{No manuf}}\right)\right)$ where "No manuf" refers to the normalizing counterfactual where we include all changes in the rest of the economy (including aggregate labor supply changes), but keep manufacturing fixed (that is, we fix all primitives in manufacturing at the initial year, keep the distribution of manufacturing plants and potential entrants fixed over time and only allow labor demand to vary with the endogenous wage over time). And "k, No manuf" is the same counterfactual as "No manuf" except that we include only changes in the rest of the economy that are also considered in counterfactual k. For looking at the role of initial conditions, $y_{2015}^{k,\text{No manuf}} = y_{1979}^{k,\text{No manuf}}$. Further details are reported in the Appendix.

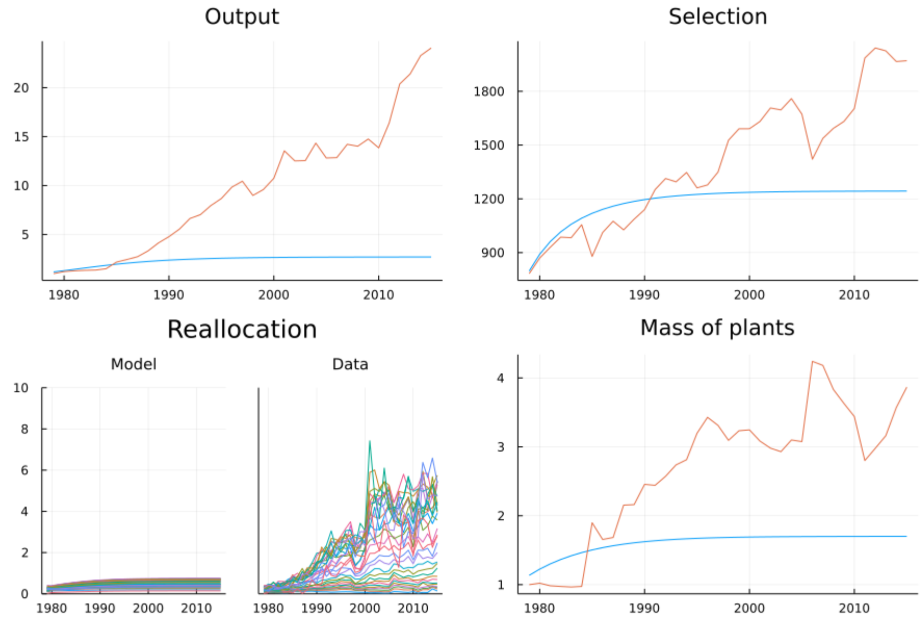

Figure 14: Results for counterfactual where economy evolves only based on initial conditions (all primitives fixed to initial level). Model results in blue and baseline data in orange. All graphs report results for manufacturing only. Top-left panel reports output, with initial year normalized to 1. Top-right panel reports average idiosyncratic productivity. Bottom-left panel reports total labor in each productivity state. Bottom-right panel reports the mass of plants.

increases steadily despite there not being any increase in the economy-wide supply of inputs (labor). This increase is driven by a slow reallocation of labor from the rest of the economy (the manufacturing employment share more than doubles from 3.5% to 8.5%) as well as a reallocation of labor across manufacturing plants. The initial plant size distribution features many small plants and few large plants, and as plants demand more labor, the right tail of the plant size distribution slowly fills up over time. Based on the structural model, we find that initial conditions explain 40% of the increase in the average size of plants over time. This is sizable but lower than predicted by the reduced-form exercise in Section 2 because the structural model takes into account that growing labor demand raises wages and slows down further labor demand.

Average plant productivity increases by roughly 40% over time, most of which is realized in the first 10 years. This important selection effect is driven by plants moving closer to their stationary distribution of productivity as well as the exit of less productive and the slow entrance of more productive plants. The growth in labor demand that leads to changes in the plant size distribution over time and the slow reallocation of labor from the rest of the economy drives a secular rise in wages, leading to a further positive selection effect as less productive plants are now even more likely to exit and less likely to enter. The increase in wages also endogenously affects misallocation as measured by the covariance between plant size and productivity.

However, most selection and average plant size effects play out over a time span of roughly 10-15 years, while the rise in manufacturing output and the reallocation of labor from the rest of the economy towards manufacturing persists much longer. The reason is the huge importance of entry dynamics in our case. Based on the initial distribution of potential entrants, the model predicts a sizable and persistent flow of new entering plants that outnumber exit, slowly driving up the mass of manufacturing plants over time. Furthermore, the accompanying rise in wages is not strong enough to discourage entry sufficiently to markedly slow down this rise in entry. This important role for entry seems to be in line with other recent evidence that points towards an outsized role for entry dynamics in explaining periods of fast growth Kim, Lee, and Shin (2021) as well as recently observed slowdowns in growth and productivity (e.g. Alon et al. 2018; Engbom 2019; Liang, Wang, and Lazear 2018).

The continuing importance of initial conditions

While previously highlighted transition dynamics may take decades to play out, an economy with a given fixed set of primitives will eventually reach its steady state. Even so, initial conditions continue to play a role in our framework. The reason is that while initial conditions induce transition dynamics towards a new steady state, aggregate changes in the meantime may shift the steady state itself. One can think of this economy (or dynamic system) as a race between transition dynamics and aggregate shocks that change the primitives of the underlying economy. Initial conditions do not become less important over time if aggregate changes are frequent and large enough such that they shift the stationary distribution faster than transition dynamics can reach the previous steady state. This novel dynamic can only be studied with a model that features transition dynamics and aggregate changes in primitives and we find strong quantitative evidence for the importance of this dynamic in our data.

To show this, we redo the previous exercise, but instead look at transition dynamics as implied by primitives of the Indonesian economy in 2015, 40 years later than before. That is, we compute a counterfactual growth path where we fix the initial distribution of plants and primitives in 2015 that are given by the potential entrant distribution, frictions, wedges, aggregate technology in manufacturing and the rest of the economy as well as aggregate labor supply. We find that initial conditions in 2015 are - by many standards - as important in 2015 than they were in 1979. Figure 15 gives details on the implied transition dynamics from 2015 until 2051. Manufacturing output almost doubles over the next 40 years, smaller than the more than doubling implied by initial conditions in 1979, but much larger in terms of levels. Based on the previous metric, the growth from initial conditions in 2015 would account for more than 50% of growth in output per worker over time. However, the drivers of growth have changed. Selection is now contributing negatively to growth as

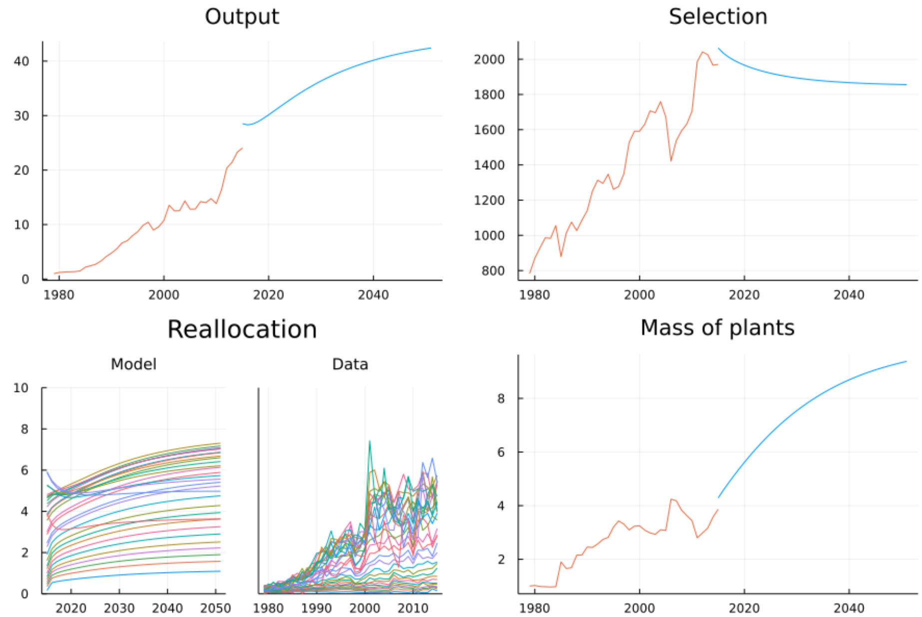

Figure 15: Results for counterfactual where economy evolves only based on initial conditions in 2015 (all primitives fixed to initial level). Model results in blue and previous data in orange. All graphs report results for manufacturing only. Top-left panel reports output, with initial year normalized to 1. Top-right panel reports average idiosyncratic productivity. Bottom-left panel reports total labor in each productivity state. Bottom-right panel reports the mass of plants.

average plant productivity regresses back to the mean of our estimated productivity process.[23] Similar to before, most of the growth is driven by an expansion of entry as implied by observed entry behavior in the last years of the data.

The role of aggregate technology

To provide some perspective on the importance of initial conditions and to relate our results to the literature, we also quantify the role of aggregate technology in manufacturing. We find that aggregate technology growth explains around 61% of the growth in output per worker over time. We quantify this driver by considering a counterfactual in which we shut down aggregate technology growth in manufacturing. Results for this exercise are reported in Figure 16. While aggregate manufacturing output is less than half in this economy, there are also about half as many plants due to a relative increase in exit and reduction in entry.

This effect is close to the direct effect of technology as captured in the per capita version of the growth accounting framework, because there are two competing effects that balance each other. On the one hand, better technology drives up wages, increases entry and reallocates labor from the rest of the economy to

[23]This regression to the mean likely captures a shortcoming of our model as we do not allow improvements in the productivity process over time nor that younger plants draw from a different productivity process.

manufacturing where it is more productive. On the other hand, aggregate technology growth also has a negative effect on the selection of plants as even low productive plants enter or find it beneficial to continue to produce. This can be seen from the Selection panel in Figure 16, where shutting down aggregate technology growth actually has a beneficial effect on the average productivity of plants.

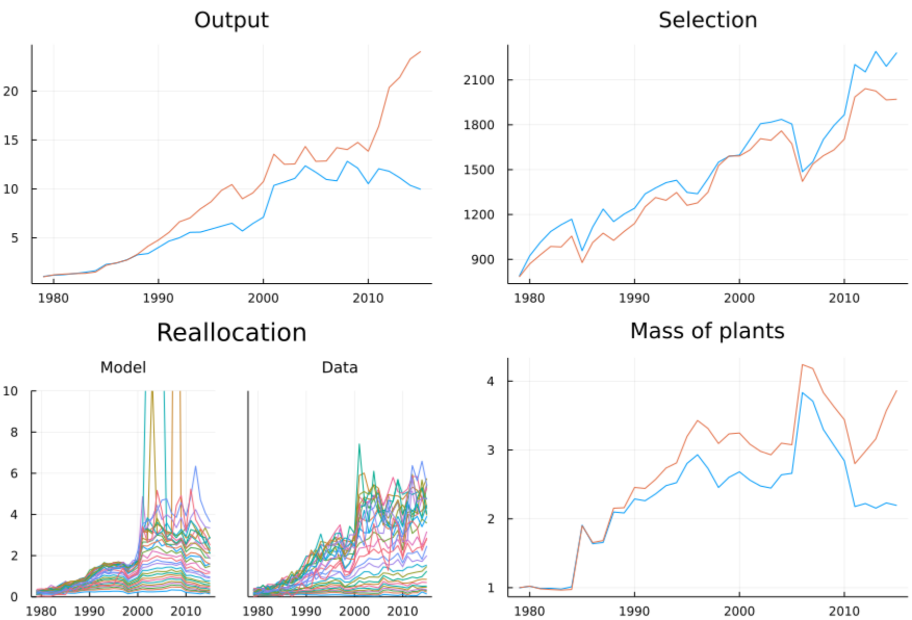

Figure 16: Results for counterfactual where aggregate technology is constant throughout. Model results in blue and baseline data in orange. All graphs report results for manufacturing only. Top-left panel reports output in the data versus model, with initial year normalized to 1. Top-right panel reports average idiosyncratic productivity. Bottom-left panel reports total labor in each productivity state. Bottom-right panel reports the mass of plants.

The role of changes in frictions

At last, we find that changes in frictions, if anything, contribute negatively to economic growth. We base this conclusion on two pieces of evidence. First, estimated changes in frictions as measured by time-varying aggregate labor market wedges reduce aggregate economic output by roughly 8% over the period 1979-2015. We find this based on a counterfactual in which we keep labor and financial market frictions at their 1979 values, shut down the role of wedges and let all other primitives move as in the baseline distorted economy. Figure 17 illustrates the evolution of the economy without changes in frictions.

The second piece of evidence is to look at plant-level wedges in the tradition of Hsieh and Klenow (2009). We find evidence for an increase in the dispersion in plant-level wedges over time, which - as in Hsieh and Klenow (2009) - depresses output and efficiency due to decreasing returns to scale in production. To show this, we compute plant-level wedges τ_{it} as: $h_{it}^{data} = (1+\tau_{it}^{k}(k))h_{it}^{model,k}$ where $h_{it}^{model,k}$ gives the model-implied optimal labor demand (for $k \in \{\text{unconstrained}, \text{constrained by aggregate wedge}\}$). Figure 18 shows an increase in

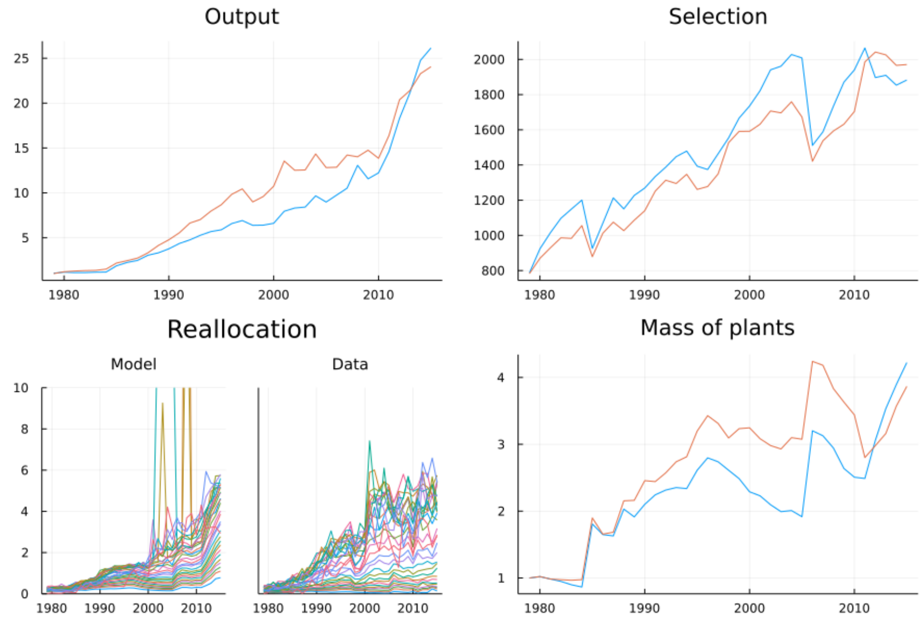

Figure 17: Results for counterfactual where wedges (and hence changes in labor and financial market frictions) are shut down. Model results in blue and baseline data in orange. All graphs report results for manufacturing only. Top-left panel reports output in the data versus model, with initial year normalized to 1. Top-right panel reports average idiosyncratic productivity. Bottom-left panel reports total labor in each productivity state. Bottom-right panel reports the mass of plants.

the dispersion of wedges over time, aggregating wedges across plants by weighting by observed labor demand. The benefit of this exercise is that in contrast to the aggregate labor market wedge, plant-level wedges do not require our model to correctly predict changes in the distribution of plants. Furthermore, in contrast to Hsieh and Klenow (2009), these wedges are consistent with dynamic input choices away from the steady state. This difference is important, because changes in wedges following Hsieh and Klenow (2009) might have simply picked up endogenous changes in the distribution of plants implied by transition dynamics and wrongly concluded that average wedges and their dispersion declined over time.

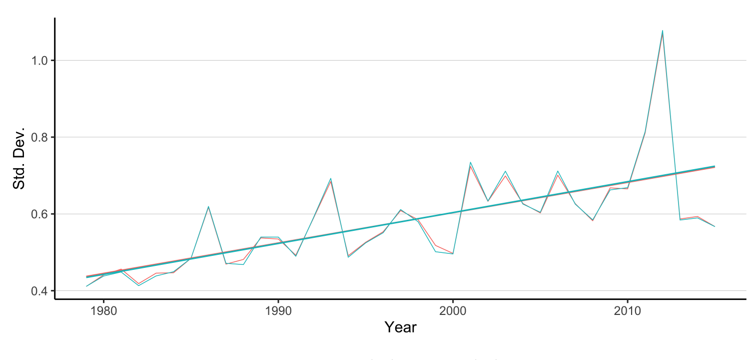

Figure 18: Yearly standard deviation of plant-specific wedges over time. Plant-specific wedges are computed as differences between observed plant-specific labor demand in the data and model-implied optimal labor demand. The standard deviation is computed by weighting wedges by observed labor demand.

In the Appendix, we show more details on estimated plant-level wedges. Specifically, we show how these wedges are informative about potential model misspecification by looking at how estimated wedges correlate over time within plants and how they correlate with further observable heterogeneity such as industries, ownership structures and age. We find that estimated wedges are broadly in line with treating them as random noise, the approach we implicitly take in this paper.

5 Robustness & Extensions

We have worked on a number of extensions, including one with constrained rational expectations. These results will be added soon.

5.1 Constrained rational expectations and perfect foresight equilibria

5.2 Industry heterogeneity

6 Conclusion

This paper has shown how plant dynamics drive economic growth over the course of development. By focusing on evidence from Indonesian manufacturing, we showed how observed economic growth was driven by a combination of technology growth, initial conditions and slow transition dynamics. The initial plant size distribution of Indonesian manufacturing in 1979 entailed the seeds of future economic growth. This is because the initial distribution featured few large plants and many medium-sized plants with a large growth potential. Output growth over the 40-year period we look at was driven by technology growth and by general input growth due to these medium-sized plants slowly wanting to grow large and the corresponding reallocation of economic resources - mostly from other parts of the economy - towards these plants.

At last, our results are in line with evidence of slow life-cycle growth of plants in developing countries (Hsieh and Klenow 2014) and the lack of large plants in developing countries (Hsieh and Olken 2014). The structural model allows us to view these stylized facts through the lens of the model and quantify their contribution to aggregate growth. Being able to study plant-level behavior in a structural model with flexible transition dynamics and growth offers the opportunity for studying in more detail the impact of a variety of policies and distortions and taking their study closer to the data.

7 References

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

Alon, Titan, David Berger, Robert Dent, and Benjamin Pugsley. 2018. "Older and Slower: The Startup Deficit's Lasting Effects on Aggregate Productivity Growth." *Journal of Monetary Economics* 93: 68–85.

Alvarez-Cuadrado, Francisco, and Mihaela I. Pintea. 2009. "A Quantitative Exploration of the Golden Age of European Growth." *Journal of Economic Dynamics and Control* 33 (7): 1437–50.

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. "Dynamic Inputs and Resource (Mis) Allocation." *Journal of Political Economy* 122 (5): 1013–63.

Asturias, Jose, Sewon Hur, Timothy J. Kehoe, and Kim J. Ruhl. forthcoming. "Firm Entry and Exit and Aggregate Growth." *American Economic Journal: Macroeconomics*, forthcoming. https://doi.org/10.1257/mac.20200376.

Baqaee, David Rezza, and Emmanuel Farhi. 2020. "Productivity and Misallocation in General Equilibrium." *The Quarterly Journal of Economics* 135 (1): 105–63.

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

———. 2017. "Productivity Growth and Capital Flows: The Dynamics of Reforms." *American Economic Journal: Macroeconomics* 9 (3): 147–85.

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.

Chen, Kaiji, Ayşe İmrohoroğlu, and Selahattin İmrohoroğlu. 2006. "The Japanese Saving Rate." *American Economic Review* 96 (5): 1850–58.

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.

David, Joel M., and Venky Venkateswaran. 2019. "The Sources of Capital Misallocation." *American Economic Review* 109 (7): 2531–67.

Engbom, Niklas. 2019. "Firm and Worker Dynamics in an Aging Labor Market." Federal Reserve Bank of

Minneapolis Minneapolis, MN.

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. https://doi.org/10.1093/qje/qjx024.

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

Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica: Journal of the Econometric Society*, 1127–50.

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

Hsieh, Chang-Tai. 2002. "What Explains the Industrial Revolution in East Asia? Evidence from the Factor Markets." *American Economic Review* 92 (3): 502–26.

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.

Kim, Minho, Munseob Lee, and Yongseok Shin. 2021. "The Plant-Level View of an Industrial Policy: The Korean Heavy Industry Drive of 1973." National Bureau of Economic Research.

Krusell, Per, and Anthony A. Smith. 1998. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy* 106 (5): 867–96.

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

Lu, Shu-Shiuan. 2012. "East Asian Growth Experience Revisited from the Perspective of a Neoclassical Model." *Review of Economic Dynamics* 15 (3): 359–76.

Melitz, Marc J., and Sašo Polanec. 2015. "Dynamic Olley-Pakes Productivity Decomposition with Entry and Exit." *The Rand Journal of Economics* 46 (2): 362–75.

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. https://doi.org/10.1257/aer.104.10.3186.

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

Poschke, Markus. 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.

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

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. 2011. "Growing Like China." *American Economic Review* 101 (1): 196–233.

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.

Young, Alwyn. 1995. "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience." *The Quarterly Journal of Economics* 110 (3): 641–80.

A Appendix

A.1 Details and robustness on decomposition results

In this section of the Appendix, we provide details on the decomposition exercise and a number of further exercises that test the robustness of the main results.

A.1.1 Formal derivation of main accounting identity

We can start by giving a formal derivation of the main accounting identity.

$$
\begin{aligned}
Y_t &\equiv \sum_i y_{it} \\
&= \sum_i z_t s_{it} f(x_{it}) = \sum_i z_t s_{it} f(x_{it}) \frac{\sum_i f(x_{it})}{\sum_i f(x_{it})} \\
&= z_t * \sum_i f(x_{it}) * \sum_i s_{it} \frac{f(x_{it})}{\sum_i f(x_{it})} \\
&= z_t * \sum_i f(x_{it}) * \sum_i (s_{it} - \bar{s}_t + \bar{s}_t) \left(\frac{f(x_{it})}{\sum_i f(x_{it})} - \frac{1}{N_t} + \frac{1}{N_t} \right) \\
&= z_t * \sum_i f(x_{it}) * \left[\bar{s}_t + N_t cov \left(s_{it}, \frac{f(x_{it})}{\sum_i f(x_{it})} \right) \right]
\end{aligned}
$$

$$
ln(Y_t) = ln(z_t) + ln \left(\sum_i f(x_{it}) \right) + ln \left(\bar{s}_t + N_t cov \left(s_{it}, \frac{f(x_{it})}{\sum_i f(x_{it})} \right) \right)
$$

A.1.2 Detailed year-to-year results

For the case of $f(x_{it}) = h_{it}^{\theta}$, we can provide detailed year-to-year results. Specifically, Figure 19 plots four main terms over time: output, input, selection and the covariance term. The technology part and the evolution of the number of plants over time are already reported in the main paper.

A.1.3 Robustness with respect to output elasticity θ

As one key robustness exercise, we consider the importance of the output elasticity θ. Given that this parameter is the main input in the decomposition exercise and determines the relative role of productivity

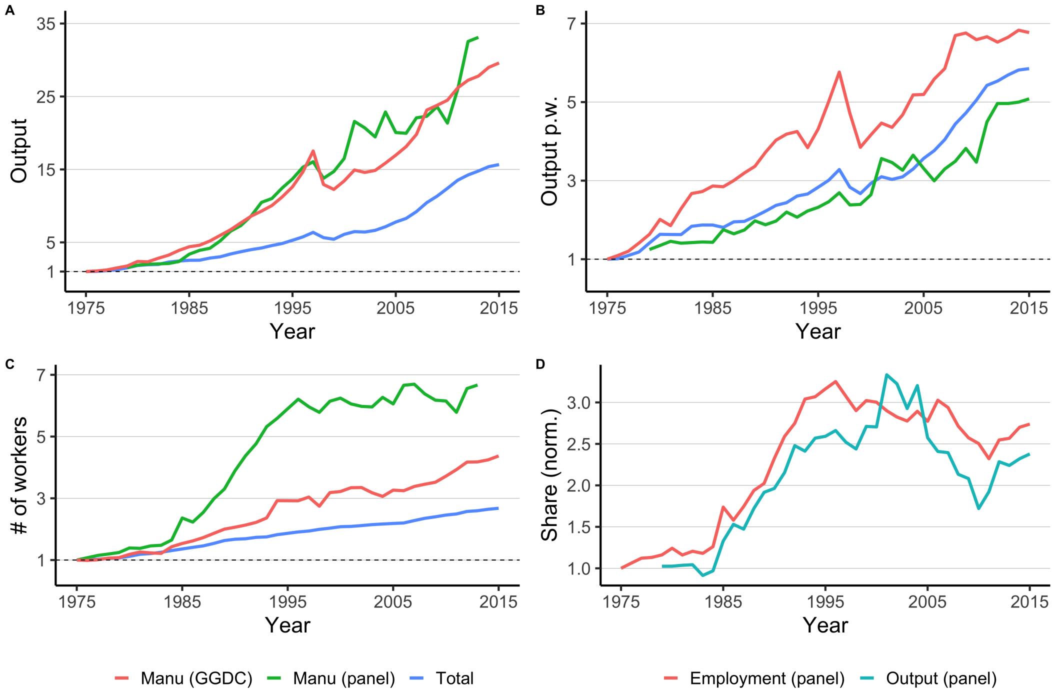

Figure 19: Evolution of four main terms in the main growth accounting identity.

versus inputs, one might expect that the decomposition results heavily depend on θ. However, not only do we have a good estimator for θ, it also turns out that the main results are surprisingly little affected by the actual choice of θ. Specifically, we consider a grid of values for θ between 0.4 and 0.9, nesting all commonly used estimates in the literature. For each value of θ, we repeat our identification strategy of aggregate technology.

(add graph!!)

A.1.4 Decomposition results for entry and exit

(Need to add these results)

A.2 Details and robustness for aggregate technology estimates

A.3 Details and robustness for production function estimation

In this section of the Appendix, we explain in more details our estimator for the production function elasticity θ. We discuss the generality of the estimator, how we estimate θ in practice as well as how we estimate standard errors. The key idea behind the estimator comes from looking at the policy functions implied by a model with adjustment costs as shown in Panel A of Figure 10. Each separate line gives the policy function for a specific idiosyncratic productivity realization in period t. Each dotted line gives the static unconstrained optimal choice of labor, which is independent of previous labor. Note that the dotted line in combination with the 45 degree line (which captures whether plant inputs grow or decline) separates the policy space into four regions. Importantly, optimal model-implied policies only happen to be in two of the four regions. In case a plant is above their optimal static unconstrained labor demand (as judged by current productivity and past labor demand), a plant always wants to weakly decrease their labor demand, while in the opposite case, a plant wants to weakly increase labor demand. The estimator exploits this geometric result by finding the θ that maximizes correctly classifying plants into these two regions and minimizing the number of plants that are classified into being in one of the other two regions.

We conjecture that the geometric result of optimal policies being concentrated entirely in the bottom left and top right region of Figure 10 holds more generally for many dynamic frictions. With flexible adjustment costs in labor, this holds for any set of economically reasonable parameters where the exact parameters only govern where plants should locate within these two regions. Similarly, the same geometric classification holds for financial frictions that prevent plants from choosing the unconstrained optimal input choice.

More formally, plants produce with:

$$y_{it} = TFP_{it} * l_{it}^{\theta}$$

They observe TFP_{it} before choosing labor demand l_{it}, but face flexible adjustment costs on labor demand l_{it} that depend on past labor demand l_{it-1} and potentially also other variables such as TFP_{it}. For any given θ, one can back out TFP_{it} using observed y_{it} and l_{it}. Furthermore, define the optimal unconstrained labor demand by

$$l_{it}^{*} = \left(\frac{\theta \mathrm{TFP}_{it}}{w_t}\right)^{\frac{1}{1-\theta}}$$

where w_t captures observed real wages. Define the object $X_{it} \equiv l_{it}^{*} - l_{it-1}$. Then based on the economic model, plants will want to weakly increase their labor demand if $X_{it} > 0$ and weakly decrease their labor demand if $X_{it} < 0$. We can denote these indicator functions by $I_{it}^{*}(\theta)$ and $D_{it}^{*}(\theta)$ and their data counterparts by I_{it}^{D} and D_{it}^{D}. To find θ, the estimator minimizes the distance between indicators that are implied by the model and observed in the data. Specifically, the loss function we use is:

$$\text{Loss}(\theta) \equiv \sum_i \sum_t \omega_{it} \left[\mathbb{1}_{I_{it}^{D}=1} (I_{it}^{D} - I_{it}^{*}(\theta))^2 + \mathbb{1}_{D_{it}^{D}=1} (D_{it}^{D} - D_{it}^{*}(\theta))^2 \right]$$

where

$$\omega_{it} \equiv \frac{(|l_{it} - l_{it-1}|)^{\varphi}}{\sum_{it}(|l_{it} - l_{it-1}|)^{\varphi}}$$

To actually estimate θ, we take into account a number of additional constraints that complicate estimation of θ in practice and force us to make certain sample restrictions and assumptions. Based on our most preferred estimate, $\hat{\theta} \approx 0.58$. We visualize the robustness of our estimated θ with respect to these assumptions in Panels A-D in Figure 20. In each panel, we estimate θ over a grid by varying one assumption while keeping the other ones fixed at our most preferred choice. Unless otherwise noted, grey dotted lines denote the choices we make.

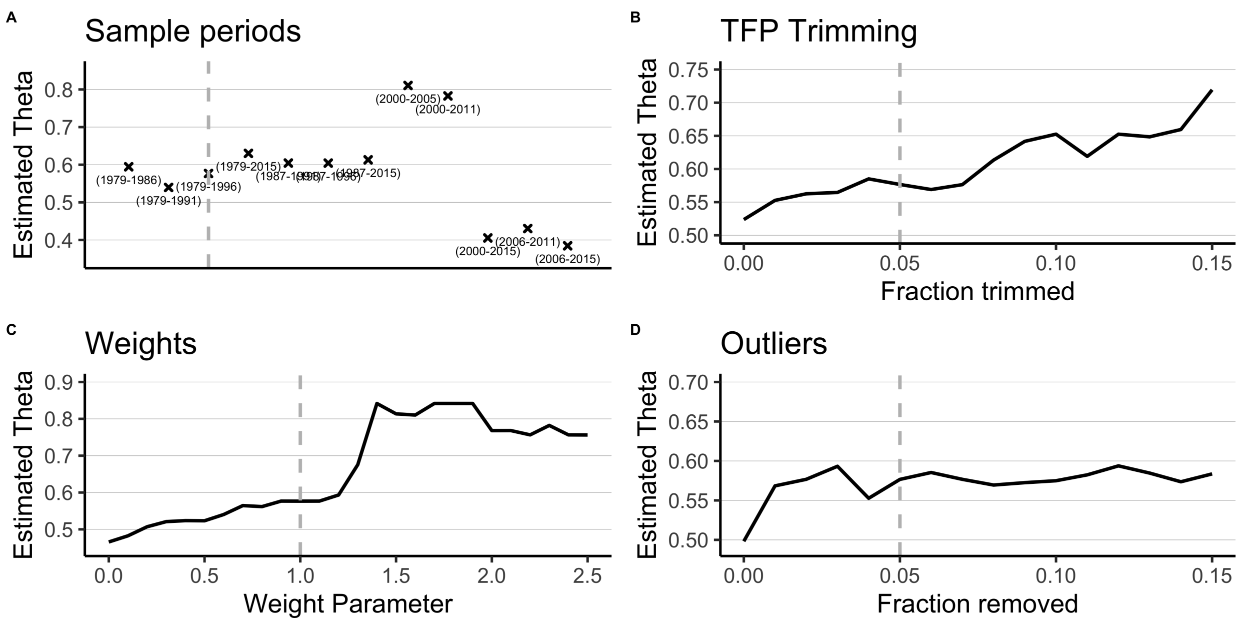

Figure 20: Robustness of production function parameter estimation. Panels A-D each vary one margin while fixing the other margins at the most preferred choice. Unless otherwise noted, grey dotted lines denote the choices we make for each margin. Panel A: Gives estimated elasticities for different sample periods (where time periods are reported in brackets below). Panel B: Considers the case of trimming extreme tails of the TFP distribution (conditional on the elasticity). Panel C: Varies the weight parameter for the weights as defined in the loss function. Panel D: Additionally considers dropping outliers by trimming a fraction of observations with the highest weights. Weights are well-defined for plants that are shrinking or growing so that such trimming will remove plants of both types.

The first complication is that the geometric classification can fail in case plants have strong expectations about better or worse future productivity or cost realizations. In the case of labor adjustment costs, plants may anticipate that their future idiosyncratic productivity will deteriorate compared to their current one, in which case they might want to decrease their labor demand even if they should increase their labor demand based on the previous discussion. This leads to misclassification. This issue plays a role in case plants are at the upper end of the idiosyncratic productivity distribution and there is strong regression to the mean for productivity that is also anticipated by plants. Similarly, plants with low productivity may anticipate future increases in productivity and thus increase their labor demand nonetheless. This means that the issue is decreasing in the persistence of the productivity process; this issue vanishes in the limit where idiosyncratic productivity is completely persistent (unit root). By a similar argument, expectations about changes in future wages or aggregate technology can also lead to misclassification. Hence, the estimator is likely to work much better in environments where aggregate technology and wages are close to constant (and this is also anticipated by plants) and where idiosyncratic productivity shows high persistence. Fortunately, measured real wages are fairly constant and estimated aggregate technology (also for different values of θ) is highly persistent (see also the results in the later part of the Appendix). There is a trade-off between choosing a time period in which real wages and aggregate technology are very stable and in choosing a long time period to maximize the number of observations. In the end, we chose as our most preferred setting the entire time period before the Asian Financial Crisis: 1979-1996. Panel A of Figure 20 varies the time-period for which θ is estimated. Estimates of θ are fairly similar for any time period before the Asian Financial Crisis, but differ when using mostly data for after the year 2000. The much lower estimates when including data from after 2006 could be caused by the issue of misclassification, by structural breaks in the production function parameters or outliers (as discussed further below).

Panel B of Figure 20 additionally trims the estimation sample by dropping x% of the top and bottom by productivity. Estimates are much more stable with respect to this change and are increasing in the amount of trimming. The reason is that plants that are dropped are more likely to be misclassified as growing their size (while they shrink in the data). We choose 5% trimming here as an intermediate or slightly conservative value.

Another complication is model-implied inaction. As shown in Figure 10, adjustment costs with fixed costs lead to inaction where plants keep labor demand constant. However, the importance of plants that are at or close to the inaction region will be determined by the weights as governed by the parameter φ. For positive φ, plants that are closer to the inaction region will be weighted less strong in the estimation of θ than plants that are further away. In Panel C, we vary φ and find that estimated θ also varies with φ. We choose the natural linear weight $\varphi = 1$, which seems to give an intermediate to slightly conservative value for θ. Additionally, we also considered dropping all plants in a specific distance to the inaction region and found that this leaves estimates for θ almost untouched. The reason is that with $\varphi = 1$, those plants will have very small weights anyway. For our most preferred estimate, we thus do not additionally drop any plants close to the inaction region. At last, we additionally drop outliers in terms of weights. The estimated θ might be disproportionately affected by a few plants who have very high weights. To obtain outlier-robust estimates, we drop x% of the plants with the highest weights (which captures both plants who grow and shrink strongly) and show results for varying x in Panel D. Again, we find stable estimates for θ and choose a value of 5% which gives an intermediate value for θ.

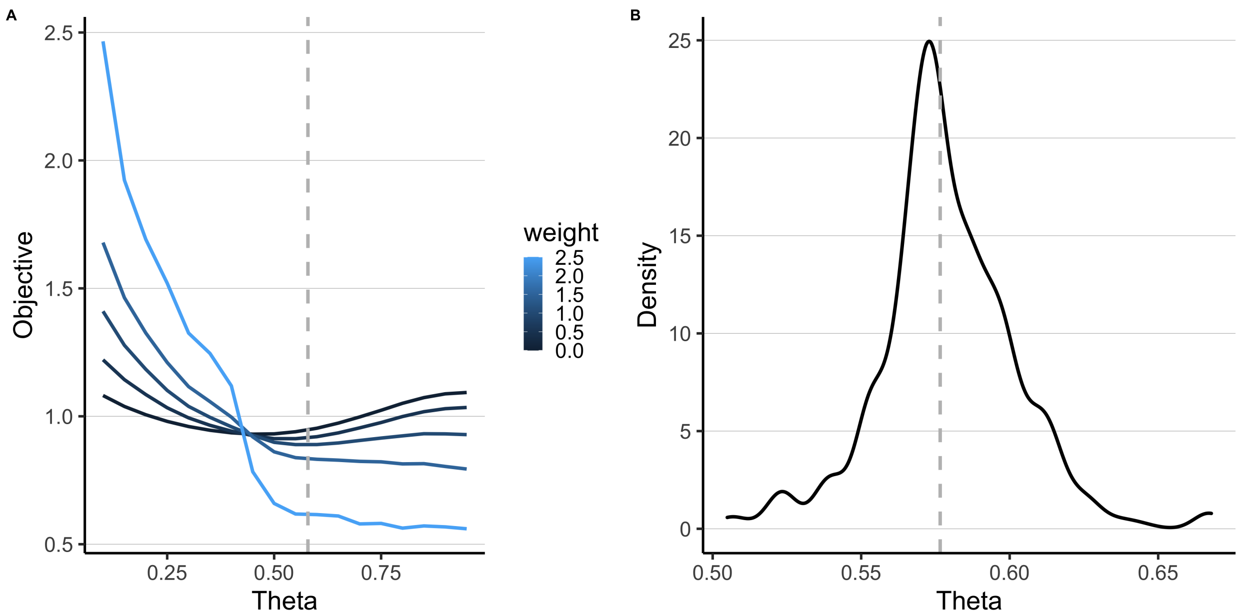

Figure 21: Visualizing identification of production function elasticity Theta. Panel A plots objective functions for different weights. The minimum of each objective function gives the respective estimate. The grey line gives the most preferred point-estimate. Panel B shows the distribution of bootstrap estimates. The grey line, again, denotes the point estimate.

In the end, our most preferred estimate is based on the time period 1979-1996, weights based on $\varphi = 1$, 5% TFP trimming and 5% of outlier trimming. The final estimation sample includes roughly 170k plant-year observations, or about 25% of all plant-year observations in the data. To quantify statistical uncertainty, we form 95% bootstrap confidence bands using the standard non-parametric bootstrap. This gives non-symmetric confidence bands that range from 0.52 to 0.626. To show identification, Figure 21 reports two different results. Panel A formally shows identification by plotting normalized objective functions over values of θ for different weights φ. For each φ, the objective function has a unique minimum, proving identification. Furthermore, the problem is smooth and well-behaved. Panel B reports the bootstrap distribution to show the adequacy of using the nonparametric bootstrap for inference. While showing general validity of the bootstrap procedure is beyond the scope of this paper, we note that this is a smooth problem (as shown in Panel A), suggesting that bootstrapping should work well here. Looking at the distribution of bootstrap estimates of θ suggests that the distribution has well-behaved tails (based on 500 bootstrap estimates, all estimates fall between 0.5 to 0.67) and is unimodal. Based on the estimated θ, one important point needs to be highlighted: $\hat{\theta} = 0.58$ is considerably larger than the average labor share in the estimation sample (around 0.46). Through the lens of the theoretical model, this means the estimator captures a situation in which plants are systematically below their optimal plant size.

To end this section, we have also alternatively considered rewriting the estimator as a Maximum Likelihood estimator. Given the binary nature of model-predictions being entirely based on the sign of X_{it}, one could consider a Probit estimator where plants observe X_{it}, but make a decision based on $X_{it} + \varepsilon_{it}$. However, given the scaling of X_{it}, this is not a well-defined problem as long as all plants draw from the same distribution of shocks. Alternatively, one could consider rescaling X_{it} by focusing for example on growth rates. However, this gives too small weights to large plants and too large weights to small plants. Estimates based on the Maximum Likelihood procedure turned out to give unrealistic parameter estimates ($\theta < 0.05$), which led us to discard the Maximum Likelihood estimates.

A.4 Details and robustness for main reduced-form exercise

(Need better intro of this section)

While the first empirical exercise given by the lines "1975 Hypothetical" is explained in the main text, the other two empirical exercises are not fully explained. These exercises account for entry and exit. To begin with, note that entry and exit is potentially very important, especially if entering plants differ from exiting plants. Of the roughly 6,800 plants 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 3 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, as can be seen from Figure (?), 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 "Hypothetical 1975 (EE)" (where EE stands for entry and exit). 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, 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 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 "Hypothetical 1975 (EE growth)". 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 long-run effects on the evolution of the size distribution.

We have robustness results for:

- varying initial distributions and transition matrices
- averaging transition matrices over multiple years
- varying assumptions on entry and exit behavior
- same exercise for real value added output and the labor wage bill

We will include these results in a later version of the paper.

A.5 Further results for plant-specific wedges

Here, we report further results on our plant-specific wedges. Note that wedges, by definition, capture differences between model-predicted labor demand and observed labor demand at the plant-level over time. There are two main interpretations of these wedges. According to the first view - the way the misallocation literature has interpreted them - our model is correctly specified and the wedges simply capture random frictions and shocks at the plant-level. They still have to be orthogonal to our main model mechanisms, however, because otherwise estimated parameters will be biased/inconsistent. An alternative view interprets them as model failure: errors that represent our ignorance of the world. In both cases, wedges should be random noise, or at least uncorrelated with other main model mechanisms. Fortunately, we can test this, revealing potential sources of misspecification and highlighting dimensions on which our and future models can be improved.

To this end, we check whether we can predict individual plant-year-specific wedges and which unmodelled aspects of the data are most predictive of wedges. In summary, we find that wedges are hard to predict and that there are few observables in the data that predict estimated wedges well. To establish this result, we start by considering unmodelled heterogeneity across plants. For example, one might expect that variation across industries is important in explaining differential labor demand patterns in the data and to the extent that the model misses this industry heterogeneity, model errors are correlated within industries. We find little evidence for such a systematic correlation of model errors across industries or any other time-unvarying plant characteristics over time.

We start by considering plant-specific fixed effects that capture any time-unvarying plant characteristics (such as industry, general ownership, location and any unobserved fixed plant types). Plant fixed effects only capture about 26% of the variance in plant-specific wedges (an upper bound given the variance of plant-specific fixed effects).[24]

Next, we verify whether wedges are autocorrelated within plants over time. We find an autocorrelation coefficient that is small and negative (around -0.2), highlighting limited negative persistence in wedges. The explanatory power of past wedges is also low, being below 5% of the variance. The natural interpretation of this is that it is not that there are specific plant-types who have consistently high or low wedges (fixed unobserved heterogeneity that we miss), but rather that the data features limited lumpiness in hiring and firing that our model misses, leading to large labor changes in one period and little changes in the next. Our model completely abstracts from unobserved shocks that affect lumpiness in investments.

[24]Throughout, we report results for $log(1 + \tau_{it})$ as this provides a more linear comparison between positive and negative wedges.

At last, among observed plant characteristics, both fixed and time-varying, we can ask which characteristics are most predictive. We find industry, district, province and year fixed effects to explain only around 2% of the variance in estimated wedges, pointing away from location- or industry-specific heterogeneity as the main culprit for model errors. For time-varying plant characteristics, we test the predictive power of plant age, ownership shares by foreigners and the government as well as plant size measured by the number of workers. We find these time-varying variables to explain little to no variation in estimated wedges. While some of these variables end up being statistically significant, their joint explanatory power does not exceed 3% of the variance in wedges (which also holds after additionally controlling for a number of time-unvarying fixed effects).

A.6 Further results on model identification, estimation and validation

A.6.1 Identification proof for θ

In this part, we simulate from the model and show that our estimator reveals θ. This is currently still work-in-progress.

A.6.2 Further validation: Life-cycle dynamics of plants

In this part of the Appendix, we show that our model can account for slow life-cycle growth of plants (Hsieh and Klenow 2014) and the lack of large plants in developing countries (Hsieh and Olken 2014). This is currently still work-in-progress.

A.6.3 Further validation results

Add further validation graphs.