Heterogeneous Mark-Ups and Endogenous Misallocation

Michael Peters*
Massachusetts Institute of Technology

October 30, 2011

Abstract

Aggregate total factor productivity depends both on firms' physical productivity (technology) and on the efficiency of the resource allocation. The empirical regularity that cross-sectional productivity differences are more pronounced in underdeveloped economies is often interpreted as evidence for the importance of misallocation to explain the cross-country variation in TFP. This paper argues, that this finding is also informative about differences in technology. I construct a simple endogenous growth model, where the cross-sectional distribution of productivity and the growth rate of aggregate TFP are jointly determined. As output markets are imperfectly competitive, the cross-sectional productivity dispersion will reflect the distribution of mark-ups. Mark-ups, however, not only cause static misallocation, but they also determine innovation and entry incentives and hence the equilibrium growth rate. If equilibrium entry is intense (for example because entry barriers are low), product markets are competitive, the productivity dispersion across firms is small and the aggregate growth rate is high. The cross-country variation in productivity dispersion might therefore be a symptom of fundamental differences in the innovation environment. Using firm-level data from Indonesia, I present both reduced form evidence for this mechanism and estimate the models' structural parameters. A policy, which reduces existing entry barriers by 10%, will increase welfare by 5%. While 10% of these gains are attributed to a reduction of static misallocation, 60% stem from a change in the equilibrium growth rate.

JEL Codes: O11, O33, O43, D42

Keywords: Endogenous mark-ups, Entry, Entry costs, Firm-level distortions, Measuring misallocation, TFP differences, Economic growth, Technological progress

*Massachusetts Institute of Technology, Department of Economics, 50 Memorial Drive, E52-391, Cambridge, MA 02142, eMail: mipeters@mit.edu. I am especially grateful to my advisors Daron Acemoglu, Abhijit Banerjee and Rob Townsend for their invaluable guidance, encouragement and support. I also thank Ufuk Akcigit, Isaiah Andrews, Thomas Chaney, Maya Eden, Roberto Fattal Jaef, Felipe Iachan, Chang-Tai Hsieh, Daniel Keniston, Amir Kermani, Sam Kortum, Benjamin Moll, Plamen Nenov, Ezra Oberfield, Michael Powell and particularly Marios Angeletos, Arnaud Costinot and Iván Werning as well as seminar participants at MIT, Chicago, the RED Mini-Conference on "Misallocation and Productivity" and the SED. I am also grateful to Ida Fariana from the Badan Pusat Statistik (BPS-Statistics Indonesia) for her cooperation in obtaining the Indonesian micro data, to Monica Martinez-Bravo for providing me with me PODES data and for the financial assistance of the George and Obie Shultz fund to acquire the data necessary for this project. All errors are my own.

1 Introduction

There is a broad consensus in the economic profession that cross-country income differences are to a large degree driven by differences in aggregate productivity (TFP).[1] Aggregative growth theory has traditionally interpreted these disparities as stemming from technological differences across countries. In the last couple of years, a new literature emerged that tends to ignore technological differences and instead focuses on the importance of misallocation of resources across firms within countries to explain the cross-country variation in aggregate TFP. Lately, these contributions have attracted substantial attention, because recent work developed an empirical accounting framework, which allows researchers to measure misallocation using readily available firm-level data. Hsieh and Klenow (2009), following the methodology of Foster, Haltiwanger, and Syverson (2008), show that in an important class of macroeconomic models, firms' productivity is proportional to their marginal products.[2][3] The dispersion of productivity across firms has therefore been interpreted as a key measure of allocative efficiency. And indeed, there is substantial evidence that poor countries are characterized by larger cross-sectional productivity differences, which is consistent with the view that differences in the static efficiency of the resource allocation are important to understand aggregate differences in TFP (Hsieh and Klenow, 2009; Bartelsman, Haltiwanger, and Scarpetta, 2009; Collard-Wexler, Asker, and De Loecker, 2011).

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

More specifically, I consider an economy, where imperfect competition on output markets endogenously generates heterogeneous mark-ups across firms. In such an environment, measured productivity differences across producers fully reflect the cross-sectional distribution of prices. These monopolistic distortions cause resources to be statically misallocated, because production factors are not allocated towards the most efficient units but rather to the ones whose monopolistic power is relatively low. From a dynamic point of view, however, these endogenous distortions govern firms' innovation incentives and hence determine the growth rate of aggregate TFP. I study an environment where productivity growth can either be generated by process innovations of current producers or stem from new firms entering the market. While incumbent firms spend innovation resources to increase their mark-up, entering firms reduce mark-ups as they adopt

[1] I will review the related literature below.

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

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

current best-practice techniques and thereby tighten product market competition. A greater equilibrium entry rate is therefore associated with a lower cross-sectional variance of mark-ups and hence less dispersion in measured productivity. Additionally, a higher equilibrium entry rate (usually) also causes the growth rate of aggregate TFP to be higher. Countries with substantial barriers to entry (e.g. license requirements or imperfect markets for venture capital) will therefore both have a high dispersion of productivity as existing firms' monopoly power is unchecked and a low rate of aggregate productivity growth. Conversely, countries that offer a level playing field for new firms will see little productivity dispersion, as entering firms keep monopoly power limited and grow at a faster rate. Hence, higher productivity dispersion not only implies a lower degree of static allocative efficiency, but is also a symptom of a more fundamental difference in the innovation environment, which has direct implications for technological differences across countries.

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

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

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

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

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

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

Related Literature This paper is mostly related to the literature trying to understand regional productivity differences (Banerjee and Duflo, 2005; Caselli, 2005; Klenow and Rodriguez-Clare, 1997). While a large literature aimed to explain these differences in the context of an aggregate production function,[4] recently various contributions focused explicitly on the behavior of individual firms and their interaction. On the one hand, Hsieh and Klenow (2009) and Restuccia and Rogerson (2008) argue that misallocation of resources across individual firms can have sizable negative effects on aggregate TFP but are agnostic about the reasons for this misallocation. On the other hand, there are papers taking a more structural perspective in that they explicitly model the microstructure of the economy. Acemoglu, Antras, and Helpman (2007) for example argue that contractual incompleteness reduces the incentives to invest into technology and Lagos (2006) shows that in an economy with search frictions, aggregate TFP depends on the primitives of the underlying search technology. Among others, Buera, Kaboski, and Shin (2011), Midrigan and Xu (2010), Jeong and Townsend (2007), Banerjee and Moll (2010) and Moll (2010) consider dynamic models of entrepreneurship and study how capital market imperfections affect aggregate productivity by distorting the allocation of talent across occupations and the allocation of capital across firms. This paper tries to build a bridge between these two approaches. While my model is structural in that firm-level distortions are generated endogenously, the reduced form of my model looks exactly like the economy considered in Hsieh and Klenow (2009) or Restuccia and Rogerson (2008). Hsieh and Klenow (2011) and Fattal Jaef (2011) also stress the dynamic interactions between productivity differences and entry, albeit in a very different way. While they analyze entry incentives given a particular set of distortions firms will eventually face, this paper claims that the causality may go the other way: the equilibrium entry rate will shape the mark-up distribution in the economy as it determines the toughness of competition.

There is also a related literature in the field of international trade stressing the importance of mark-ups. Bernard, Eaton, Jensen, and Kortum (2003) allow for mark-up heterogeneity precisely to account for the positive correlation between measured productivity, firm efficiency and export status, and Atkeson and Burstein (2008) find that variable mark-ups are important to account for the observed pattern of producer and consumer prices. Epifani and Gancia (2011) argue that non-constant mark-ups might be

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

important to assess the welfare effects of trade liberalization - if trade opening reduces mark-ups in sectors whose mark-up was already below average, welfare might decrease as the higher dispersion in mark-ups reduces TFP. Edmond, Midrigan, and Xu (2011) on the other hand provide quantitative evidence from a calibration exercise using plant-level data that the pro-competitive effect of trade-liberalization, whereby the entry of foreign producers decreases mark-ups, induces welfare gains, which are an order of magnitude larger than the usual love-for-variety effects of trade opening. Finally, the importance to distinguish physical productivity from prices has been stressed in the recent empirical literature on how trade liberalization affects firms' productivity. De Loecker (2011) shows that controlling for unobserved prices reduces the estimated productivity increase substantially. Relatedly, De Loecker and Warzynski (forthcoming) find that firms increase their mark-ups upon entry into exporting and argue that mark-ups might be responsible for the large productivity premium of exporting firms usually found in firm-level data. Finally, the distinction between technological progress, aggregate productivity growth and the role of mark-ups has a long tradition in macroeconomics. See for example Basu and Fernald (2002).

In terms of modelling choices, I formalize the evolution of productivity as a Schumpeterian quality ladder model of vertical innovation as in Aghion and Howitt (1992) and Grossman and Helpman (1991). The characterization of the balanced growth path shares some similarities to Klette and Kortum (2004) and Lentz and Mortensen (2008), although I do not consider multi-product firms in my framework. To generate heterogeneous mark-ups, I take an otherwise standard CES demand system, but drop the assumption of different firms producing differentiated varieties.[5] In contrast, different producers engage in Bertrand competition so that prices depend on firms' relative productivities (see also Bernard, Eaton, Jensen, and Kortum (2003) and Acemoglu and Akcigit (forthcoming)).

The structure of the paper is as follows. In the next section I present the model. First, I characterize the static allocations and show that the impact of mark-ups on the aggregate economy is summarized by two sufficient statistics. I then turn to the endogenous determination of equilibrium mark-ups. As mark-ups depend only on the cross-sectional distribution of physical productivity, this requires a model of how relative technologies evolve. I consider a framework with both entry and productivity improvements by incumbent firms and show that the endogenous stationary distribution of mark-ups is Pareto, where the shape parameter is given by the equilibrium entry intensity. Section three is devoted to the empirical analysis, which proceeds in two steps. In the reduced form analysis, I first assess to what extent different aspects of the productivity data are consistent with my theory of mark-ups and compare the predictions to a simple model with credit constraints. Then I provide evidence that the regional entry intensity is negatively correlated with both the level of mark-ups and their dispersion. Finally, I estimate the structural parameters of the model using GMM, compare the implied mark-up distribution to the firm-level data and consider the welfare implications of a reduction of entry costs in different regions of Indonesia. There I also provide a simple extension of the model, where I explicitly allow for trade between the different regions in Indonesia. Section 4 concludes and an Appendix gathers some of the mathematical details.

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

2 The Model

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

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

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

$$Y\left(t\right) = exp\left(\int_0^1 ln\left(\sum_{j\varepsilon S(\nu,t)} y_j\left(\nu,t\right)\right) d\nu\right), \tag{2}$$

where $y_j(\nu,t)$ is the quantity of variety ν bought from producer j and $S(\nu,t)$ denotes the number of firms active in sector ν at time t. Hence, different varieties ν and ν' are imperfect substitutes, whereas there is perfect substitutability between different brands within a variety. The market for intermediate goods is monopolistically competitive, so that firms take aggregate prices as given but compete a la Bertrand with producers offering the same variety. The production of intermediates is conducted by heterogeneous firms and requires capital and labor, both of which are hired on frictionless spot-markets. The only source of heterogeneity across firms is their factor-neutral productivity. In particular, a firm producing variety ν with current productivity[6] q produces output according to

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

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

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

It is useful to characterize the static and dynamic allocations separately. By static allocations, I refer to the equilibrium allocations for a given level of aggregate capital $K(t)$ and a given number of firms and their productivities $\left[S(\nu,t), q_j(\nu,t)\right]$. By construction these allocations neither depend on the inter-

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

temporal preferences of households, nor on the details of the innovation environment. This static analysis will show that the cross-sectional distribution of mark-ups is fully determined from the distribution of physical productivities across firms and characterizes how different moments of this distribution affect the aggregate economy via their effects on mark-ups. To derive the (long-run) equilibrium distribution of mark-ups I therefore need to characterize the evolution of physical productivities. This is done in the dynamic analysis, where I solve for firms' innovation incentives and the equilibrium degree of entry.

2.1 Static Allocations: Heterogeneous Mark-ups and Misallocation

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

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

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

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

where $q_F(\nu,t)$ is the follower's productivity.[7] The equilibrium mark-up is therefore given by

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

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

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

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

and the producer's factor demands as

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

Finally, the producing firm's profits are given by

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

The expressions above show that the cross-sectional variation in profits and factor demands can be entirely traced back to firm-specific mark-ups. Furthermore, it is precisely these varying mark-ups, which induce an inefficient allocation of resources across plants and hence have detrimental effects on aggregate TFP. This is most clearly seen from (8), which shows that producers' factor demands depend on their mark-up and not on their physical efficiency.

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

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

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

To characterize the general equilibrium of this economy, let $L_P(t)$ be the number of workers in the production sector.[8] Aggregate output can then be written as

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

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

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

Hence, at the aggregate level, this economy looks exactly like the canonical neoclassical growth model where aggregate TFP is determined both by physical productivity $Q(t)$ and the monopolistic mark-ups

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

summarized in the TFP distortion index $M(t)$. Furthermore, it can be easily verified, that static first-best aggregate output is given by $Y^{FB}(t) = Q(t) K(t)^{\alpha} L^{1-\alpha}$, so that $Y(t) = Y^{FB}(t) M(t)$. $M(t)$ therefore measures precisely the static efficiency losses of monopolistic pricing. Mark-ups also manifest themselves in equilibrium factor prices, as these are given by

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

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

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

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

is the *factor price distortion index.* This term stresses that monopolistic pricing reduces equilibrium factor prices for two reasons.[9] Not only is aggregate TFP (and hence each production factors' marginal product) lower compared to the competitive economy, but even conditional on aggregate TFP, equilibrium factor prices are below their social marginal products as $\frac{R(t)}{MPK(t)} = \frac{w(t)}{MPL(t)} = \Lambda(t)$.[10] Hence, $M(t)$ and $\Lambda(t)$ are sufficient statistics for the static TFP and factor price consequences of monopolistic power. As both $M(t)$ and $\Lambda(t)$ are fully determined from the distribution of mark-ups, Proposition 1 characterizes how this distribution affects aggregate TFP and factor prices in this static economy, i.e. the economy where the distribution of leading productivities $[q(\nu,t)]_{\nu=0}^{1}$ and the level of capital $K(t)$ is given.

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

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

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

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

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

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

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

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

□

Proposition 1 shows that the different moments of the mark-up distribution have very different impacts on the economy. In particular, factor prices are entirely insensitive with respect to a higher dispersion of the underlying mark-up distribution - they only depend on the mean. A shift in the level of mark-ups will therefore reduce factor prices and increase equilibrium profits. For the case of aggregate TFP, exactly the opposite is true: whereas a higher dispersion of mark-ups will show up as a lower TFP for the aggregate economy, TFP is not affected by level shifts. Intuitively, aggregate TFP is determined by the allocation of resources across firms. If all mark-ups increase by a constant proportion, relative prices will be unaffected and hence still provide the appropriate signals about firms' relative productivity. Note that the canonical case of constant mark-ups as generated by a CES demand system with differentiated products is a special case of Proposition 1: TFP will be identical to its efficient competitive counterpart but monopolistic power reduces factor prices.[11]

2.2 Dynamics: The Equilibrium Distribution of Mark-Ups

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

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

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

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

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

In this economy innovations can stem from two sources: either current producers can experience technological improvements, or new firms can enter the market and thereby replace the current quality leader. If a producer with current productivity q experiences an innovation, it reaches the next step of the quality ladder so its new productivity is given by λq. For the case of entry, suppose that leading technologies are common knowledge for the process of innovation so that the entrant in sector ν enters the market with productivity $\lambda q(\nu,t)$, where $q(\nu,t)$ is the leading quality in sector ν. This formulation is appealing in the current context, because it focuses on the different allocational consequences of the two sources of productivity improvements. While both entrants and incumbents increase the frontier technology by the same amount, the implications for equilibrium mark-ups and allocational efficiency are very different. In case the innovation stems from the current producer of variety ν with a quality advantage of $\Delta(\nu,t)$, the equilibrium mark-up for that variety *increases* by a factor λ. However, when productivity growth is induced by entry, the equilibrium mark-up for variety ν *decreases* by a factor $\lambda^{\Delta(\nu,t)-1}$, as the new entrant is only a single step ahead on the quality ladder.[12]

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

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

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

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

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

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

defined by

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

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

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

i.e. is exponential with parameter $ln(1+x)$. This also implies that the stationary distribution of mark-ups is Pareto with shape $\frac{ln(1+x)}{ln(\lambda)}$ as[13]

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

This result is similar to Bernard, Eaton, Jensen, and Kortum (2003), who also generate a (truncated) Pareto distribution of mark-ups in their model. The crucial difference is that here the shape parameter is endogenous as it is determined from the equilibrium entry intensity. If entry is intense, the shape parameter is large so that mark-up heterogeneity and the average mark-up declines.[14] Figure 1 depicts the density of mark-ups for two different values of the entry intensities, which correspond to the inter-quartile range of estimated entry intensities across regions of Indonesia (see section 3.3 below). The high-entry environment is characterized by less mark-up dispersion in the cross-section of firms and the average mark-up is lower. Finally, the two distortion indices summarizing the impact of mark-ups on the aggregate economy are also fully determined and given by

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

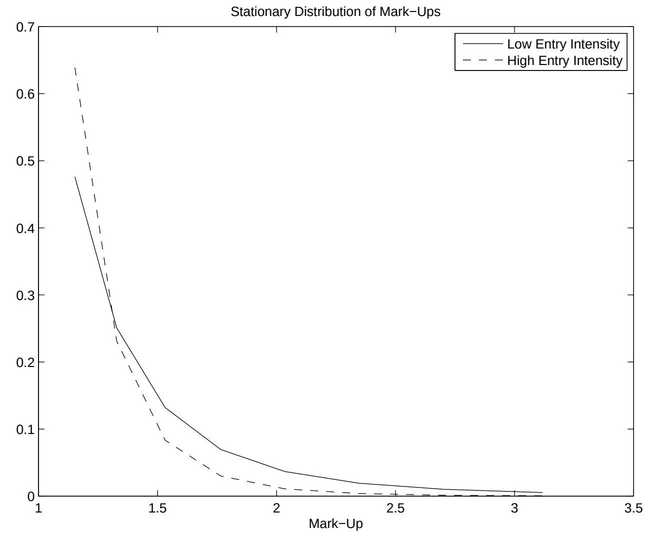
$$M(x) = \frac{exp\left(E\left[ln\left(\xi^{-1}\right)\right]\right)}{E\left[\xi^{-1}\right]} = \frac{\lambda^{-E[\Delta]}}{E\left[\lambda^{-\Delta}\right]} = \lambda^{-\frac{x+1}{x}} \frac{(\lambda-1)+\lambda x}{x}. \tag{23}$$

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

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

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

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



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

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

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

2. *increases aggregate TFP and aggregate income, as $M(x)$ is increasing in x*[15]

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

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

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

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

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

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

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

The specific functional form of the stationary mark-up distribution contained in (21) relies of course on the particular entry process. As entrants always enter with the minimal mark-up in the economy, it is intuitive that a higher entry rate will reduce both the level of mark-ups and their dispersion. Here I want to briefly discuss a different assumption concerning the entry process, which suggest that the results in Proposition 2 are more general. Suppose, as before, that incumbent firms always climb one rung of the ladder but that entrants' innovations are larger in that they enter with a blueprint of quality $q(\nu,t)\lambda^b$, where $b > 1$. In that environment, the stationary distribution of mark-ups will be exactly the same as before, but just with a higher minimum. In particular, F_ξ will be given by (21), where now m takes the values $\lambda^b, \lambda^{b+1}, ...$ Hence, all the results derived above hold true in that environment. The key is of course the stationary of the distribution. As incumbent firm will always weakly increase their mark-up, the measure of firms charging mark-ups less than λ^b will strictly decrease over time. In the stationary distribution, λ^b is again the minimum mark-up in the economy, which is charged by entering firms. Hence, entry has a pro-competitive effect.[16]

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

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

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

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

conditional on the distribution of productivity and capital simply by affecting the stationary distribution of mark-ups and with it allocational efficiency.

2.3 Equilibrium Innovation, Entry and Balanced Growth

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

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

where φ parametrizes the productivity of the innovation technology and $\gamma > 1$ ensures that the cost function for innovation is convex so that there is a unique solution. The term $\lambda^{-\Delta}$ implies that innovations are easier the bigger the productivity advantage Δ and is similar in spirit to the assumption of knowledge capital made in Klette and Kortum (2004) or the setup in Atkeson and Burstein (2010). In particular, this assumption is necessary to make the model consistent with a balanced growth path and with Gibrat's Law, i.e. with the fact the firms' growth rates are independent of size (Sutton, 1997; Luttmer, 2010).[17]

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

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

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

- *prices* $[p(\nu, t|q, \Delta)]$ *and factor demands* $[k(\nu, t|q, \Delta), l(\nu, t|q, \Delta)]$ *are consistent with firms' static profit maximization*

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

- *aggregate consumption* $[C(t)]$ *is consistent with consumers maximizing utility*
- *entry rates* $[z(\nu, t|q, \Delta)]$ *are consistent with free entry*
- *innovation rates* $[I(\nu, t|q, \Delta)]$ *are optimal given the value functions* $[V(\nu, t|q, \Delta)]$
- *the value functions are equal to the expected present discounted value of profits*
- *the evolution of* $[Q(t), M(t), \Lambda(t)]$ *is consistent with the entry and innovation rates*
- *prices* $[r(t), w(t)]$ *are consistent with market clearing, where* $r(t) = R(t) - \delta$

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

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

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

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

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

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

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

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

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

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

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

as individuals can either chose to work, earning the wage rate $w(t)$, or they can try to enter with an improved technology. In the latter case, they are successful with a flow rate of η and earn a net value (after paying the entry costs) of $V(1,t)-\chi w(t)$ as they always enter with a productivity advantage of $\Delta=1$.[18] This is also the reason why equilibrium entry rates are constant across different varieties.

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

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

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

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

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

As shown in the Appendix, (31) and (32) not only have a unique solution, but they also imply straightforward comparative static results. In particular, it can be shown that the entry intensity x is decreasing

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

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

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

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

and the two distortion indices $M(x)$ and $\Lambda(x)$ are constant. Given these equilibrium values, the rest of the economy behaves exactly like the neoclassical growth model with exogenous productivity growth - the only difference being that both factor prices and aggregate TFP are affected by $M(x)$ and $\Lambda(x)$, which act like a tax from the agents' point of view. In particular, in the framework of Chari, Kehoe, and McGrattan (2007), $\Lambda(x)$ is exactly the investment and labor wedge and $M(x)$ the efficiency wedge. This is seen from the budget constraint of the representative household[19]

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

where $MPL(t)$ and $MPK(t)$ are the marginal products of labor and capital (which agents take as parametric), L_P is the equilibrium number of production workers and $\Pi(t) = (1 - \Lambda(t)) Y(t)$ are the aggregate profits in the economy. Hence, (34) shows that from the consumers' point of view, mark-ups are just like a tax $\Lambda(x)$ on their factor supplies which are rebated back via lump-sum transfers (through $\Pi(t)$). Also, the equilibrium marginal products $MPL(t)$ $(MPK(t))$ are lower than the first-best benchmark by the efficiency wedge $M(t)$ (see (13) and (14)).

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

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

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

As clearly seen from (35), holding the growth rate fixed, a higher entry intensity x increases the BGP level of capital per production worker as the equilibrium interest rate R will be higher. This is entirely due to the pro-competitive effect of entry: by reducing mark-ups, the demand of capital will be high, which induces capital accumulation. (36) then shows that entry affects consumption even conditional on the level of capital $\tilde{k}$. As $M(x)$ is increasing in the entry-intensity, allocative efficiency will increase. Hence, the

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

composition of productivity growth between entrants and current firms affects economic welfare through four margins: it affects the growth rate g_Q, it governs the dynamic incentives to accumulate capital via $\Lambda(x)\,M(x)$, it determines static allocative efficiency through $M(x)$ and it controls the allocation of labor.

Proposition 4. *Consider the economy described above. There exists a unique balanced growth path equilibrium. The equilibrium growth rate is given by* $g_Y = \frac{1}{1-\alpha} g_Q$*, where* g_Q *is given in* (33) *and* (x, I) *are uniquely determined from* (31) *and* (32)*. The steady-state levels of normalized capital* $\tilde{k}$ *and consumption* $\tilde{c}$ *are given in* (35) *and* (36)*.*

2.4 Static Misallocation and Dynamic Incentives and Aggregate Productivity Differences

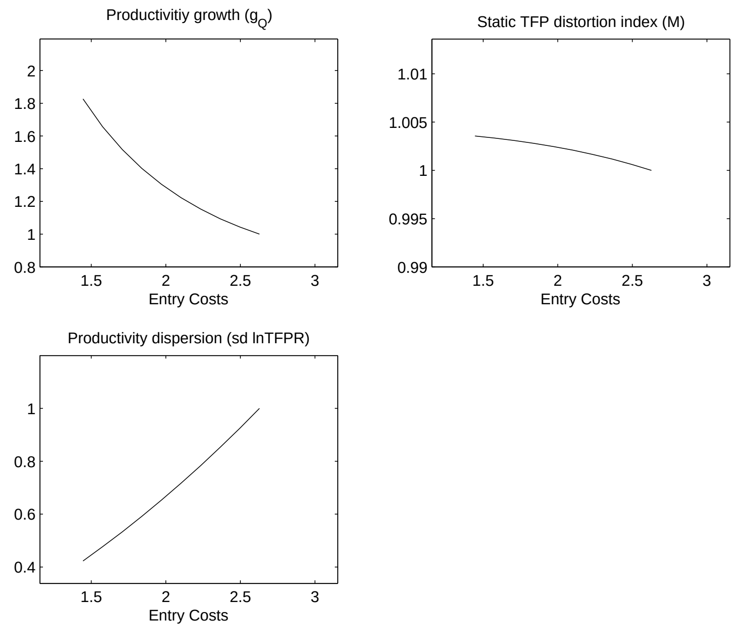
According to this paper, the static degree of misallocation and the aggregate productivity growth rate are equilibrium outcomes that are jointly determined. To see this most clearly, note that (using (21) and (33)) we can express the equilibrium distribution of mark-ups (and hence the distribution of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

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

That the cross-sectional dispersion of marginal products[20] and the equilibrium growth rate are related via (37), is particularly important when we try to explain the cross-country variation in aggregate TFP through the cross-firm dispersion of productivity within countries. In particular, once this interdependence is taken into account, the source of variation for the observed productivity dispersion in the cross-country data is crucial. Consider for example the case of entry costs χ. In Figure 2 below, I consider the simple comparative static exercise of a reduction of the entry costs. In particular, I calculate the equilibrium outcomes (z, I) for different values of the entry costs and plot the resulting equilibrium growth rate g_Q, the TFP distortion index M and the dispersion of productivity (as measured by the standard deviation of $ln(TFPR)$) relative to the economy with the highest costs. As expected, the equilibrium growth rate and allocative efficiency increase and the dispersion of measured productivity declines, when barriers to entry are lowered.

Now suppose the world in 1900 consisted of a collection of identical economies, which only differed in their entry costs χ. In what sense would productivity differences across firms be informative about the cross-country variation in aggregate TFP? When studying the cross-section of countries in 2011, we would conclude that rich countries show less productivity dispersion than poor countries. In fact, a cross-sectional regression of aggregate TFP on the dispersion of productivity would have an R^2 of one and hence would "explain" the entire cross-country variation in TFP. But if we were to reduce misallocation by equalizing marginal products, we would be disappointed: the implied TFP gains from reallocating

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



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

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

resources are small. In fact, at the parameter values used for Figure 2 (which stem from the structural estimation conducted in Section 3.3 below), the static gains are below 1%. More important than the actual number however is the conceptual point that the standard thought experiment of decomposing aggregate TFP into "technology" ($Q(t)$) and "allocative efficiency" ($M(t)$) gets blurred once productivity growth and misallocation are jointly determined. In particular, it is tempting to think of cross-sectional productivity differences as being informative only about the "allocative efficiency"-component, precisely because revenue productivity at the firm-level does not depend on technology (or physical productivity) if marginal products were equalized. This economy is an example, where this conclusion is not warranted. Cross-sectional productivity differences are *very* informative about the "technology" part of aggregate TFP (see (37)), because both of them are equilibrium objects, which are determined simultaneously from firms' dynamic incentives.[21]

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

3 Empirical Analysis

In this section I will take the model to the data to test its main implications. The empirical analysis has two main parts. In the reduced form analysis in Section 3.2, I use the fact that firms' revenue productivity is proportional to the mark-up (see (10)). Hence, as in Hsieh and Klenow (2009), I measure firm-specific wedges but then interpret the properties of these wedges through the lens of the theory. In particular, I will first present both cross-sectional and life-cycle evidence that the empirical patterns of measured productivity are consistent with this theory but harder to reconcile with theories where firms are constrained in their input choices. Then, I test the main qualitative prediction and show that a higher degree of entry is negatively correlated with the level and the dispersion of mark-ups in a given product market (which I proxy as a geographic region). In Section 3.3 I estimate the model's structural parameters by Generalized Methods of Moments (GMM). Specifically, I will estimate region-specific entry costs (χ_r) and region-specific innovation productivities (φ_r), so that the model generates regional variation in the entry intensity (x_r) as an equilibrium outcome. As the entry intensity is a sufficient statistic for the entire distribution of mark-ups in the economy, I can then revisit the firm-level data and ask how much of the cross-sectional productivity dispersion the model is able to explain. Furthermore, I will also conduct a policy experiment, where I reduce the costs of entry, and I will decompose the welfare consequences into their static and dynamic components. While static welfare gains are fully described by the change in the TFP distortion index M, the dynamic gains stem from a change in the growth rate, a reallocation of labor across occupations and an increased capital accumulation. These two empirical approaches are not only complementary from an econometric point of view, but they also correspond to two different conceptual exercises. In the reduced form analysis I focus directly on the observed distribution of productivity and use it to infer mark-ups. For the structural estimation I will only employ moments about the dynamic properties of the economy (e.g. the rate of entry and the rate of productivity growth of incumbent firms) but do not use the information contained in the cross-sectional productivity distribution in the set of moments.

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

3.1 The Data

I will estimate the model using Indonesian plant-level data. The empirical analysis is based on two data sources, which are described in more detail in the Appendix. My main data set is the Manufacturing Survey of Large and Medium-Sized Firms (Statistik Industri), which for example has also been used in Amiti and Konings (2007) and Blalock, Gertler, and Levine (2008). The Statistik Industri is an annual census of all manufacturing firms in Indonesia with 20 or more employees, and I will use data from 1991 to 2000.[22] The final sample has about 70.000 firms.[23] Most importantly, the data contains information on the geographic region, the respective firm is located in. This allows me to calculate regional entry rates. To do so, I identify a firm as an entrant when it appears first in the data. The regional information is recorded at the level of a regency (kabupaten). There are roughly 240 regencies in Indonesia. As the manufacturing sector in Indonesia is fairly concentrated, many regencies contain too few establishments for a meaningful analysis. I therefore aggregate the data at the level of the 33 provinces.

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

3.2 Reduced Form Analysis

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

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

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

3.2.1 Productivity and Mark-Ups

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

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

To evaluate if measured productivity is consistent with (38), it is useful to compare the implications with a theory where firms are constrained in their input choices. For the sake of concreteness, I want to focus here on the case of credit-market frictions.[24] Suppose the data was generated by the following economy (which is essentially the setup of Hsieh and Klenow (2009) augmented by collateral constraints). There is a measure one of firms, each of them being the sole provider of a differentiated variety ν. The final good is the usual CES aggregate $Y(t) = \left(\int_0^1 y(\nu)^{\frac{\sigma-1}{\sigma}} d\nu\right)^{\frac{\sigma}{\sigma-1}}$ and each firm produces according to the production function $f(q,k,l) = ql^{1-\alpha}k^{\alpha}$. Capital and labor are traded on a spot market with prices (w,R) and firms face collateral constraints of the form

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

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

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

where $\overline{A}(w,R,Y,\theta)$ depends only on parameters and equilibrium quantities, which are common across firms.[25] Hence, firms are constrained if their wealth relative to their productivity q falls below some threshold. Being credit constrained has implications for measured productivity precisely because constrained firms have higher marginal products than their unconstrained counterparts. It is easy to show that $ln(TFPR)$ is given by

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

where $1[.]$ is an indicator variable. This shows that (a) constrained firms' revenue productivity is high, (b) the cross-sectional variation is generated by the variation of the single statistic $\tilde{A}$, which measures firms'

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

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

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

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

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

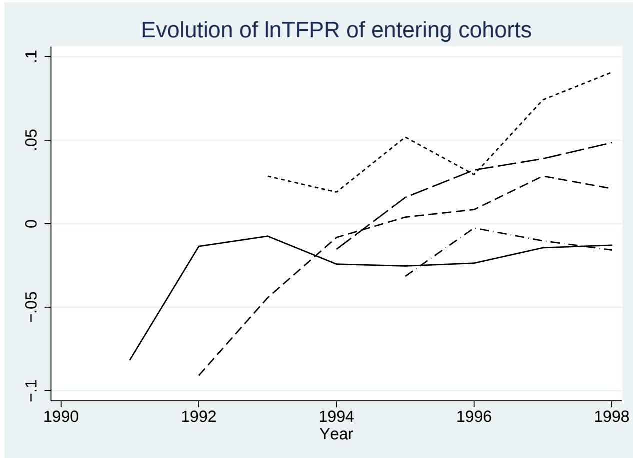
wealth relative to their productivity and (c) it is the low-$\tilde{A}$ firms that are more likely to be constrained and that have high productivity conditional on being constrained.

To assess if heterogeneous mark-ups are a potential determinant of the firm-level variation of revenue productivity, consider first Table 1, where I compare the productivity of entrants' relative to their incumbent counterparts, by regressing $ln\,(TFPR)$ on a dummy variable if the firm entered in the respective year. Consistent with this paper, entrants' productivity is 3.5% lower than the one of incumbent firms.[26] To square this finding with theories of credit-constraints (or any theory which generates productivity dispersion from constraints on inputs), one would have to argue that entering firms are relatively unconstrained. However, many theories with imperfect capital markets imply that young firms tend to be the constrained agents, that accumulate funds during their life-cycle (see for example Clementi and Hopenhayn (2006)).

It is precisely this life-cycle pattern where the implications of the different theories differ the most. According to this paper, I would expect a cohort of entering firms to increase their revenue productivity as they gradually increase their mark-up over their competitors. Simple models of credit constraints however predict the opposite. If firms are constrained when entering the market, they have incentives to accumulate funds over time so that their relative wealth $\tilde{A}$ will increase. This will reduce their revenue productivity along the life-cycle.[27] Using the panel dimension of the firm-level data I can study this dynamic pattern. In Figure 3, I depict the average (log of) revenue productivity for different cohorts entering the product market in different years. More precisely, the figure shows the evolution of log TFPR after taking out a full set of year, sector and region fixed effects. Figure 3 shows an upward trend in revenue productivity

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

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



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

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

along the life cycle. The same pattern is also reported in Hsieh and Klenow (2011), who note that for firms in India "marginal products of capital and labor (as summarized by TFPR) are also growing with age" (Hsieh and Klenow, 2011, p. 12). According to the theory of this paper, this upward trend is generated by firms increasing their mark-up through productivity improvements. For Figure 3 to be consistent with a model of credit market frictions, entering firms have to become poorer relative to their productivity as they age.

Finally, the Indonesia data allows me to directly look at the importance of credit constraints as it contains information on the source of funding for annual investment expenses. In particular, I observe if firms use FDI, if they attracted foreign loans or if they managed to acquire funds on the national capital market. I want to interpret each of these variables as identifying firms which are relatively less constrained.[28] Columns one to three of Table 2 contain the results for either of these measures. There is a very strong *positive* correlation, which is not what we would expect as unconstrained firms should have low productivity. In column four, I use the cross-section of 1996, where my dataset contains additional information from the economic census, which was added to the usual manufacturing census. Firms were asked if the lack of capital was a constraint in their expansion plans. However, column four shows that constrained firms actually have *lower* revenue productivity, which is exactly the opposite of what we expect to find if the variation in revenue productivity was generated from a simple model with credit constraints.

These results are of course not sufficient to reject one of the theories in favor of the other one and this

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

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

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

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

is also not the goal of this exercise.[29] What the results do show however, is that the general patterns we see in the data are at least consistent with a mechanism, that firms with high productivity might *not* be firms that are "unlucky" by being credit constraint or subject to stifling regulation, but that firms have high revenue productivity precisely because they were lucky enough to increase their physical productivity faster than their competitors and are rewarded for doing so with the opportunity to post a high mark-up.

3.2.2 Mark-Ups and Entry

The main qualitative prediction of the theory is that product markets, where entry is intense, should be characterized by low and compressed mark-ups. To study this regional correlation, it is useful to measure mark-ups from firm-specific labor-shares, as (8) implies that[30]

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

where τ_i is observable in the micro- data up to scale.[31] The theory implies that the distribution of mark-ups is Pareto with shape $\frac{ln(1+x)}{ln(\lambda)}$, so that

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

In Figure 4 below, I plot $ln\left(1 - F_\xi\left(m, x\right)\right)$ against $ln\left(m\right)$ and I distinguish regions according to their entry intensity. Figure 4 shows that these distributions are different. In particular, low entry regions display a thicker tail of firms charging a large mark-up. This implication is very much in line with the theory and at the heart of the mechanism of this paper: The distribution of measured productivity is more compressed in high entry environments, as new entrants keep the environment competitive by keeping prices in line with productivity improvements. Comparing Figure 4 with (43) also shows where the theory is not borne out. The theory implies that both schedules are linear and that the relationship in the entry-intense region should have a steeper slope. While it is the case that the schedule is steeper,[32] the relationship is not linear but shows some curvature.

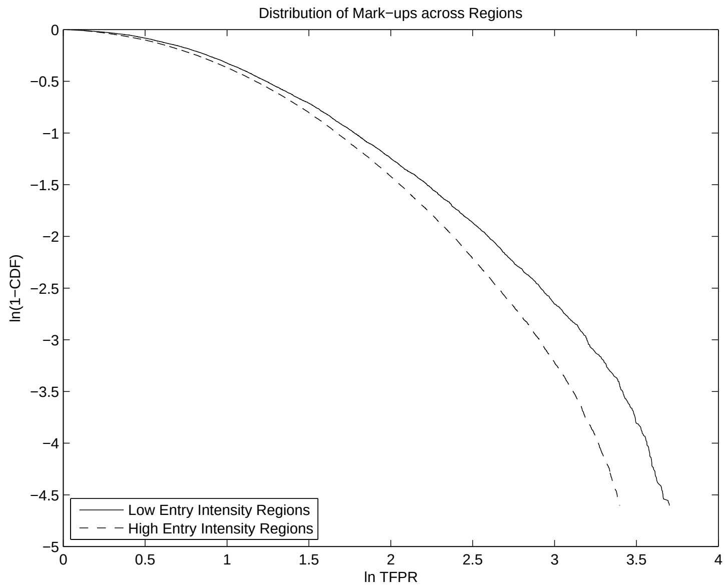
The relationship between the regional entry rate and the distribution of mark-ups depicted in Figure 4 can of course by purely driven by omitted variables that affect both the entry rate and the measured wedges in the cross-section of firms. To control for at least some of those influences, I will now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

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

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

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

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



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

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

mark-ups. Using the distribution function for log mark-ups (20), it follows that

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

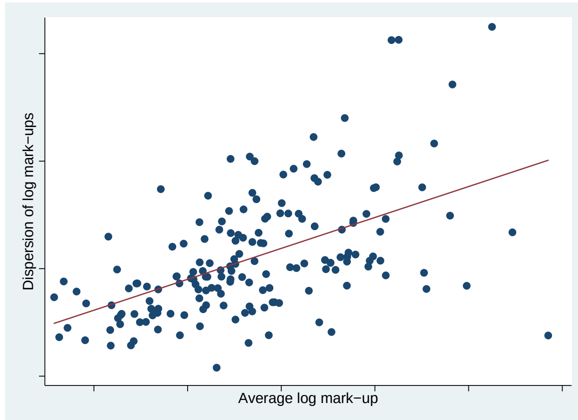
$$sd\left[ln\left(\xi\right);x\right] = ln\left(\lambda\right)\frac{\left(1+x\right)^{1/2}}{x} \quad (45)$$

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

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

The first prediction is contained in Figure 5, where I depict the simple correlation between the average mark-up and their dispersion. There is a strong positive correlation as the theory predicts.[33] Turning to

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



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

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

the second prediction, it is apparent from (44)-(46) that all these moments are decreasing in the entry intensity x. To test this prediction, I consider regressions of the form

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

where $J_{r,t}$ is one of these moments, D_t is a set of year fixed effects, $Entry_{r,t}$ is the entry rate in region r at time t, $IS_{r,t}$ controls for the industry composition in region r and T_r is a set of regional controls drawn from the regional PODES files described above.[34] The theory predicts that $\phi < 0$. The results are contained in Table 3. Columns one and two take the average log markup as the dependent variable and show that there is a negative correlation with the regional entry rate. While column one presents the simple correlation, column two shows that this correlation is neither driven by a comovement of entry and mark-ups over time nor by regional heterogeneity in their sectoral composition. In column three I include the regional control variables from the PODES files to control for some aspects of regional heterogeneity. Including these leaves the coefficient on the entry rate basically unchanged. While the size of the respective region (when measured by its population) is negatively correlated with the average mark-up, the agricultural employment share and the number of bank branches per village are positive correlated with the average mark-up. Neither the share of villages accessible by asphalted roads (as a measure of infrastructure), nor the number of informal firm seems to have a significant impact. Columns four to eight use different quantiles of the mark-up distribution as the dependent variable. For brevity I report only the specification with the entire set of controls. Again there is a strong positive relation between each of the quantiles and the entry rate, so that more intense entry indeed seems to induce first-order stochastic shifts on the

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

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

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

Table 3: Entry and the Distribution of Mark-Ups

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

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

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

fully determine the impact of mark-ups on the aggregate economy as the two distortion indices $\Lambda = E\left[\lambda^{-\Delta}\right]$ and $M = \frac{\lambda^{-E[\Delta]}}{E[\lambda^{-\Delta}]} = \frac{\Psi}{\Lambda}$ are the two sufficient statistics determining welfare.[35] According to Proposition 2, there is a monotone relation between Λ and Ψ and the entry intensity. To test this prediction, I follow the same approach as in (47) and run regressions of the form

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

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

The results are reported in Tables 4 and 5. Consider first the case of Λ, contained in Table 4. In the first column I report the simple correlation between the entry rate and $\Lambda_{r,t}$. The coefficient is positive and significant. That this relation is not purely driven by aggregate shocks over time is shown in column two, where I include a full set of year fixed effect. The coefficient hardly changes. Column three controls for the industrial share to account for the fact that some sectors might both be more prone to entry and have lower mark-ups. Like in Table 3 above, I then include the same set of regional controls to at least partially control for some regional characteristics. This does not change the partial correlation with the entry rate, which is not surprising given that none of these regional controls shows a significant correlation with the misallocation measure Λ. In column five, I use the panel structure of the data, which allows me to control for time-invariant regional characteristics by including regional fixed effects. Again this does not affect the coefficient substantially. Columns six and seven contain two additional specifications. In column six, I consider the logarithm of Λ as the dependent variable. The point estimate implies that an increase of the entry rate by one standard deviation (which is equal to 0.1), increases the factor price distortion index by 3%. While this seems very small, note that this coefficient does not have a structural interpretation, because according to the theory it is the entry intensity $x = \frac{z}{I}$, which determines the

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

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

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

Table 4: Entry and the degree of misallocation: Λ

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

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

Table 5: Entry and the degree of misallocation: Ψ

degree of misallocation. I will come back to the quantitative implications of the theory in section 3.3, when I estimate the structural parameters of the model. Finally, column seven uses a "residual" measure of misallocation. More specifically, instead of constructing Λ directly from the firm-specific labor shares observed in the data, I use the residual labor shares $\tau^{Res} = \tau - \hat{\tau}$, where $\hat{\tau}$ are the predicted values from a regression of τ on a full set of sector fixed effects and the firm-specific capital-labor ratio.[36] Hence, $\Lambda^{Res} = E\left[\tau^{Res}\right]$ is the factor price distortion index after technological heterogeneity at the firm-level is at least crudely controlled for. While the point estimate slightly lower, it is still positive and significant.

The analogous results for $\Psi_{r,t}$ are contained in Table 5. Columns one to four again show a robust positive correlation with the regional entry rate. Contrary to the case of Λ, the regional controls now have some explanatory power but do not change the partial correlation with the entry rate. As Ψ is a non-linear

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

transformation of log mark-ups, the pattern is similar to the one contained in Table 3, i.e. the population size is negatively correlated with mark-ups and both the financial density and the agricultural share is positively correlated with mark-ups. Column five includes a full set or regional fixed effects and columns six and seven again consider the log of Ψ and the residual measure of Ψ as a dependent variable.[37] The correlation is positive and significant (in column seven, the associated p-Value is just above 0.1).

While the empirical results reported above are consistent with the model, I briefly want to consider two alternative explanations. Consider first the case of capital market imperfections, in particular the simple formalization provided above. If the data was generated by that model, I would identify credit constrained firms as charging a high mark-up. Now suppose that the efficiency of the financial system (as parametrized by the collateral multiplier θ) varies across regions so that financially underdeveloped regions will see more financially constrained firms. If the regional entry rate is correlated with the regional financial development, I would conclude that entry reduces monopoly power albeit it is just the case that high entry regions are such that less firms are constrained (or firms are less constrained). As above, I will again consider firms as being relatively unconstrained whenever they are financed through FDI or raise funds on the national capital market of via foreign loans.[38] Given these measures, I investigate whether the share of unconstrained firms is indeed positively correlated with the regional entry rate, and if the entry rate has an independent effect on the degree of misallocation above and beyond regional differences in the share of unconstrained firms. The results are shown in Table 6. Columns one to three show that these measures are positively correlated with the regional entry rate, but insignificantly so. Hence, it is not surprising that columns four and five show that the entry rate is still a significant predictor of the two misallocation measures Λ and Ψ once these measures are controlled for.

The second alternative explanation I want to consider are preferential policies. In particular, the modelling device of "firm-specific taxes" used in Restuccia and Rogerson (2008) or Hsieh and Klenow (2009) is sometimes interpreted as a stand-in for actual policies like regulation, taxes or bureaucratic red tape affecting firms differentially. In particular, firms that I identify as having large monopoly power might in fact just face politically oriented barriers to expand. If additionally regions with good policies are also characterized by more dynamic entry, I would observe exactly the correlations in the data, which I interpret as being informative about the pro-competitive effects of entry. To argue that this is unlikely to be the case, I follow a similar strategy as for the case of credit constraints. In particular, I observe if either the federal or the local government has a financial stake in the firm. In Table 7, I first show that the regional entry rate is uncorrelated with the share of firms with a government stake. Columns three and four then show that the entry rate remains significantly correlated with the two distortion measures Λ and Ψ after these government ownership measures are controlled for. Note however, that the share of firms that have some government financing is also a significant corollary of an improved resource allocation.[39]

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

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

[39]Here I briefly discuss the results of various robustness checks, the details of which are available upon request. Concerning the main results reported in Tables 4 and 5, I redid the regressions using the usual robust standard errors instead of the

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

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

Table 6: The importance of credit constraints

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

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

Table 7: Government ownership

Figure 6: Map of Indonesia

3.3 Structural Estimation

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

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

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

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

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

Table 8: Economic geography of Indonesia

Identification To see which parameters of the model are identified from the microdata, consider a single region in a single time period. The model has 8 parameters $(\alpha, \rho, \delta, \varphi, \eta, \chi, \gamma, \lambda)$. As I am mostly interested in the innovation environment, I will not be concerned with estimating (α, ρ, δ) but set them exogenously at conventional levels. In particular, I will set the capital share α to 0.3, the depreciation rate δ to 0.1 and the discount rate ρ to 0.05. This leaves 5 parameters to be identified. As the model has a unique equilibrium for the innovation and entry rate, the equilibrium conditions (31) and (32) implicitly define functions

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

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

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

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

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

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

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

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

where $\tilde{\chi} = \frac{1+\eta\chi}{\eta}$. Hence, from the equilibrium conditions, I can only identify the innovation productivity φ and the net entry productivity $\tilde{\chi}$, which has a technological component η governing the efficiency of blueprint creation and an institutional component χ determining the costs of actual entry of new firm once the idea is created. Without more detailed information on where firms spend most of their resources prior to entry, it is not possible to tell η and χ apart. For the policy analysis below I will conceptually think of lowering entry barriers through a reduction of χ, but given that I can only identify $\tilde{\chi}$, this is only an interpretation.[40]

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

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

intensity x. Formally, the moment conditions are

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

where $E\left[g_{Q,r,t}\right] = E\left[\frac{Q_r(t+1)-Q_r(t)}{Q_r(t)}\right]$ and $Q_r\left(t\right)$ is calculated by aggregating the micro data. Note that these moments hold for all regions *and* all time periods. However, the model (being a balanced growth model) does not have anything to say about the time-variation in the data. Hence, I focus entirely on the regional variation and average all the moments over the time dimension. This yields $4 \times 7 = 28$ moments to estimate 16 parameters $\left(\lambda, \gamma, \left[z_r, I_r\right]^7_{r=1}\right)$. (53) and (54) then provide the equilibrium mapping to the deep structural parameters $\left[\varphi_r, \tilde{\chi}_r\right]^7_{r=1}$. While the first three moments can be directly calculated from the microdata, I do not observe detailed occupational categories in the Indonesian data. Hence, I proxy the share of innovators by the within-firm expenditure share on white-collar non-production workers.

3.3.1 Estimation results

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

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

where $\hat{\Omega}$ is the estimated variance-covariance matrix of the 28 moments and G is the matrix of moments. I estimate $\hat{\Omega}$ by bootstrapping the moments from the firm-level data.[41] The estimated parameters are contained in Table 9. The innovation step-size λ is estimated to be equal to 1.17 so that every innovation increases productivity by 17%. This is a little higher than the preferred value of 1.05 in Acemoglu and Akcigit (forthcoming) but within their range of values they consider ($\lambda \in [1.01, 1.2]$). The confidence intervals also show that λ is relatively precisely estimated. The curvature of the innovation technology is estimated to be 1.73 so that the cost function is convex as the theory requires. However, values of $\gamma < 1$ are in the confidence region. The theoretically admissible value of $\gamma = 1$ is the 12% quantile of the distribution for $\hat{\gamma}$.

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

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

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

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

Table 9: Results of GMM Estimation

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

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

Table 10: Comparing the implied moments

parameters, incumbent firms seem to enjoy a productivity advantage of generating innovations. Note however, that the confidence intervals for φ_r and χ_r are relatively large.

To see where the variation in the estimated parameter stems from, consider Table 10, where I report the moments in the data and the moments as estimated by the model. The model does a very good job to account for the pattern of entry and the share of non-production workers. This is due to the fact, that the entry rate is precisely estimated (so that the efficient GMM estimator puts a large weight on these moments) and that the share of non-production workers is the only moment that depends on γ, so that γ can basically be chosen freely to match that moment. With respect to the innovation rate of incumbent firms, the model has problems to account for the extreme outcomes in Jakarta and Kalimantan. For Jakarta, we observe a relatively small entry rate but still aggregate growth. This can only be rationalized through innovation by incumbent firms. For Kalimantan, exactly the opposite is true. Both the entry rate and the growth rate of incumbent firms is high, but aggregate productivity is at the low end. Hence, the model underestimates incumbents' innovation outcomes. Finally, with respect to the aggregate growth rate, the model under-predicts the regional variation, because of the commonality of the step size λ and the restrictions the other moments put on the entry and innovation rate. This is especially seen in the region Sulawesi, which despite its high entry rate does not see any productivity growth. With a constant step size for λ, it is impossible for the model to match these two moments simultaneously. Table 10 also conveys where the respective parameters are identified from. If we compare the regional pattern of the entry rates and the estimated entry costs $\tilde{\chi}$, it is clearly seen that regions with low entry are identified as regions where entry barriers are high. Similarly, regions where the aggregate growth rate and innovation rate of incumbent firms is estimated to be high (Sumatra or Sulawesi), are markets where the innovation technology is of high productivity. Intuitively, level differences of growth mostly identify the innovation productivity φ_r and differences in the entry rate identify the entry costs $\tilde{\chi}_r$.

3.3.2 Mark-ups and the distribution of productivity

Given the estimated parameters, the distribution of mark-ups is therefore fully determined so that I can calculate both the region-specific distortion indices Λ_r and M_r and different measures of productivity dispersion. The results from this exercise are contained in Table 11. Consider first the top panel. The median mark-up is between 10% and 20%, which is consistent with De Loecker and Warzynski (forthcoming) who estimate mark-ups between 17% and 28%. The average mark-up is higher, precisely because the theory predicts that the distribution of mark-ups is Pareto. Columns 3 and 4 contain the two distortion indices. The distortions in TFP are modest. On average, heterogeneous mark-ups reduce TFP by about 1.5-2% relative to the first best and there is little variation across regions. This is different for the case of factor prices. Column 5 shows that mark-ups reduce equilibrium factor prices by 25% relative to their social marginal product. The variation across regions is also more substantial than for the case of TFP. The reason for this pattern is contained in Proposition 1: Variation in the entry intensity affects the first moment of the resulting mark-up distribution much more than its second moment. Hence, mark-ups mostly affect factor prices. It is interesting to note that this is exactly what I found in the reduced form analysis, where I showed that the regional entry rate is negatively correlated with the average mark-up but that its relation to the dispersion of mark-ups is insignificant (see Table 3). The same result emerges again from the structural estimation despite the fact that I did not exploit any information about the cross-sectional productivity distribution in the estimation.

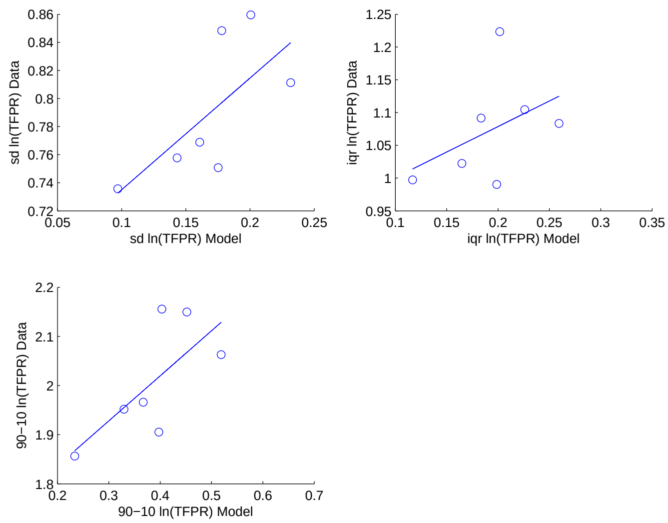
In Panels 2 to 4, I compare the implied productivity dispersion (calculated from (44)-(46)) with the one estimated from the micro-data. In the first column of the respective panel I report the estimated standard deviation, inter-quartile range and the 90-10 difference from the Indonesian firm-level data. These numbers are very similar to the one reported in Hsieh and Klenow (2009).[42] The second column reports the implied measures from the theory and third column contains their ratio, which I take as a measure of explanatory power of the theory. On average the model accounts for about 20% of the cross-sectional variation in productivity and this is roughly the same for the different measures. As the variation in the microdata is

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

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

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

Table 11: Dispersion of TFPR: Data versus Model



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

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

larger than predicted by the model, the theory does a better job in those regions where the entry intensity is low and the mark-up is high (in Kalimantan for example). The last panel finally contains an important check for the theory: Is the regional variation in the estimated entry intensity correlated with the observed productivity distribution in the microdata? The answer is affirmative: the correlation between the model's implication and the measures observed in the data is between 0.45 and 0.7. A graphical representation of this correlation is contained in Figure 7, which shows a scatter plot of the observed and implied dispersion measures across regions. There is a clear positive relation between the model's prediction and the respective moments in the data, but with only 7 regions, this is of course only suggestive. However, given that the GMM estimation only used information about the dynamic environment, there is no mechanical reason for this correlation to occur.

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

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

which again hold for all regions and all time periods. Like before, I will only use the regional variation and average out the time dimension. By estimating the parameters and studying the implied moments, it is apparent which margins of the data are putting discipline on the theory to account for both the cross-sectional dispersion and the dynamic behavior of the economy simultaneously. In Table 12, I again

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

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

Table 12: Moments using the cross-sectional information

report the comparison between the data and the model's implication at the estimated parameters. In the last two columns, I show that the model is able to account for the cross-sectional dispersion rather well. However, this comes at the expense to account for other features of the data. From (56) it is apparent that in order to generate a large cross-sectional dispersion of productivity, the theory requires the step size λ to be large and the entry intensity x to be small. These conflicting objectives are seen in Table 12. At the estimated parameters, the model underestimates the entry rate, generates too large a growth rate for incumbent firms and implies a rate of aggregate productivity growth, which exceeds the one observed in the data. Furthermore, the implied median mark-up is around 70%, which is also counterfactually high. Hence, the theoretical mechanism laid out here is not able to account for entire productivity dispersion in the cross-section of firms - according to Table 11 roughly 80% of it remains unexplained.

3.3.3 Policy and Welfare

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

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

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

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

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

The results of the policy experiment are contained in Table 13. The first row contains the size of the policy change, namely a decrease in the entry costs by 10%. This increases the entry-intensity by about 25% and reduces the dispersion of productivity by about 15%. In terms of the misallocation measures Λ and M, this increase in the entry intensity increases factor prices (relative to their marginal product) by 1% to 2.5% but has a small effect on static misallocation (as measured by M), where the increase is at most half a percentage point.[43] Columns 6 to 8 trace the effect of this policy change on the equilibrium allocation of workers across occupations. While the number of workers in the entry sector, increases by roughly 2-3%, these workers are mainly drawn from incumbent firms who see their innovation employment decline. In fact, in most regions the reduction of entry costs frees up labor resources and production employment increases. However, this is not necessarily so. In Central Java, whose estimated innovation productivity is low, the change in factor prices causes formerly employed production workers to start

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

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

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

Table 13: Effects of a reduction of barriers to entry

working in the entry sector. Finally, a reduction of the entry costs increases the equilibrium growth rate in all regions by about one or two tenth of a percentage point.

The second panel shows the decomposition contained in (58). The reduction in static misallocation raises the level of consumption by around half a percent. The Reallocation-component is larger, as more intense entry increases the number of production workers and the level of capital (except for Central Java). On average, an improved allocation of resources through a change in entry incentives increases consumption by around 2%. Finally, the Growth-component induces a slight fall in those consumption gains, because an increase in the growth rate will decrease the level of consumption as more resources are needed to have the capital stock grow at the economy-wide growth rate. The next panel contains the same decomposition for the change in the capital stock. By construction there is no "static" gain where factors are held constant. Given the changes in M and Λ (especially the latter) and the reallocation of workers, normalized capital increases by around 3-4%, with again Central Java being the exception. As for the case of consumption, the higher growth rate compensates slightly for those improvements in the allocation of resources, which reduces the normalized capital-stock.

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

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

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

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

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

3.4 Inter-regional trade

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

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

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

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

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

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

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

It is precisely (61), which will - by affecting region r's terms of trade - create the linkages across regions.

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

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

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

Consider a BGP where the (endogenous) growth rate of aggregate productivity in region r is given by $\dot{Q}_r(t) = Q_r(t) g_r$ and "global" output Y grows at rate g_Y.[44] The equilibrium return to capital in region r is still given by $R_r = \alpha\Lambda_r \frac{s(r)Y}{K_r}$, where $\Lambda_r = E_r\left[\lambda^{-\Delta}\right]$ only depends on the entry intensity in region r. Along the BGP, interest are constant (though not necessarily equal across regions) so that the *regional* capital stock grows at the same rate as *global* output, i.e. $\dot{K}_r(t) = K_r(t) g_Y$. Hence, regional (physical) output grows at rate (see (62)) $g_{Y_r} = g_r + \alpha g_Y$, i.e. productivity growth induces a region-specific component, while capital accumulation is the same across regions. From (60), $Y = \prod_r Y_r^{s(r)}$ so that

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

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

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

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

Now suppose that regional product markets in Indonesia were integrated in that way. To see that this would leave the empirical identification intact, note first that the moments regarding entry and the within-firm share of innovators are not affected by the possibility of trading, as labor markets are assumed to clear within each region. Similarly, the growth rate of incumbent firms can be identified from the

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

firm-level data, because $TFPR$ still only depends on the mark-up λ^{Δ} and not on the regional price index. The aggregate growth rate g is different. As I do not use region-specific price deflators in the estimation, the aggregate growth rate will *not* be region-specific but this model with free trade implies that a revenue-based measure of aggregate productivity should grow at the same rate across regions as terms of trade and regional productivity growth are not separately identified. Hence, allowing for free trade would only add an additional cross-region restriction that productivity should grow at the same rate. The heterogeneity in the aggregate growth rate of productivity across regions reported in Table 10 would then be interpreted as pure measurement error. However, the model is still identified in the same way as the closed-economy version above, because even there I restricted the step-size λ to be common across regions.

4 Conclusion

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

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

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

References

Acemoglu, D., and U. Akcigit (forthcoming): "Intellectual Property Rights Policy, Competition and Innovation," *Journal of the European Economic Association*.

Acemoglu, D., P. Antras, and E. Helpman (2007): "Contracts and Technology Adoption," *American Economic Review*, 97(3), 916–943.

Acemoglu, D., S. Johnson, and J. Robinson (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91(5), 1369–1401.

Acemoglu, D., and J. Ventura (2002): "The World Income Distribution," *Quarterly Journal of Economics*, 117(2), 659–694.

Aghion, P., and P. Howitt (1992): "A Model of Growth through Creative Destruction," *Econometrica*, 60(2), 323–351.

Amiti, M., and J. Konings (2007): "Trade Liberalization, Intermediate Inputs and Productivity," *American Economic Review*, 97(5), 1611–1638.

Armington, P. S. (1969): "A Theory of Demand for Products Distinguished by Place and Production," *IMF Staff Paper*, 16, 159–178.

ATKESON, A., AND A. BURSTEIN (2008): “Pricing-to-Market, Trade Costs, and International Relative Prices,” *American Economic Review*, 98(5), 1998–2031.

——— (2010): “Innovation, Firm Dynamics, and International Trade,” *Journal of Political Economy*, 118(3), 433–484.

BANERJEE, A., AND E. DUFLO (2005): “Growth Theory through the Lens of Development Economics,” in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf, vol. 1A, chap. 7, pp. 473–552. Elsevier B.V.

BANERJEE, A., AND B. MOLL (2010): “Why Does Misallocation Persist?,” *American Economic Journal: Macroeconomics*, 2(1), p. 189–206.

BARSEGHYAN, L., AND R. DICECIO (2009): “Entry Costs, Misallocation and Cross-Country Income and TFP Differences,” Working paper, Federal Reserve Bank of St. Louis.

BARTELSMAN, E. J., J. C. HALTIWANGER, AND S. SCARPETTA (2009): “Cross-Country Differences in Productivity: The Role of Allocation and Selection,” Working paper 15490, NBER.

BASU, S., AND J. G. FERNALD (2002): “Aggregate Productivity and Aggregate Technology,” *European Economic Review*, 46, 963–991.

BERNARD, A., J. EATON, J. B. JENSEN, AND S. KORTUM (2003): “Plants and Productivity in International Trade,” *American Economic Review*, 93(4), 1268–1290.

BLALOCK, G., P. J. GERTLER, AND D. I. LEVINE (2008): “Financial Constraints on Investment in an Emerging Market Crisis,” *Journal of Monetary Economics*, 55(3), 568–591.

BUERA, F., J. KABOSKI, AND Y. SHIN (2011): “Finance and Development: A Tale of Two Sectors,” *American Economic Review*, 101(5), 1964–2002.

CASELLI, F. (2005): “Accounting for Cross-Country Income Differences,” in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf, vol. 1A, chap. 9, pp. 679–741. Elsevier B.V.

CHARI, V., P. J. KEHOE, AND E. R. MCGRATTAN (2007): “Business Cycle Accounting,” *Econometrica*, 75(3), 781–836.

CLEMENTI, G. L., AND H. HOPENHAYN (2006): “A Theory of Financing Constraints and Firm Dynamics,” *The Quarterly Journal of Economics*, 121(1), 229–265.

COLLARD-WEXLER, A., J. ASKER, AND J. DE LOECKER (2011): “Productivity Volatility and the Misallocation of Resources in Developing Economies,” Discussion paper, NBER Working Paper 17175.

DE LOECKER, J. (2011): “Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity,” *Econometrica*, 79(5), 1407–1451.

De Loecker, J., and F. Warzynski (forthcoming): "Markups and Firm-Level Export Status," *American Economic Review.*

Djankov, S., R. La Porta, F. Lopez-De-Silvaes, and A. Shleifer (2002): "The Regulation of Entry," *Quarterly Journal of Economics*, 117(1), 1–37.

Edmond, C., V. Midrigan, and D. Y. Xu (2011): "Competition, Markups and the Gains from International Trade," Working paper, Federal Reserve Bank of Minneapolis.

Epifani, P., and G. Gancia (2011): "Trade, Markup Heterogeneity and Misallocations," *Journal of International Economics*, 83(1), 1–13.

Fattal Jaef, R. N. (2011): "Entry, Exit and Misallocation Frictions," Discussion paper.

Foster, J., J. C. Haltiwanger, and C. Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, 98(1), 394–425.

Gancia, G., and F. Zilibotti (2009): "Technological Change and the Wealth of Nations," *Annual Review of Economics*, 1, 93–120.

Grossman, G., and E. Helpman (1991): "Quality Ladders in the Theory of Growth," *Review of Economic Studies*, 58(1), 43–61.

Hall, R., and C. Jones (1999): "Why Do Some Countries Produce So Much More Ouput Per Worker Than Others?," *Quarterly Journal of Economics*, 114(1), 83–116.

Hopenhayn, H. (1992): "Entry, Exit and Firm Dynamics in Long-Run Equilibrium," *Econometrica*, 60(5), 1127–1150.

Hsieh, C.-T., and P. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124(2), 771–807.

——— (2011): "The Life-Cycle of Plants in India and Mexico," Discussion paper, Working Paper.

Jeong, H., and R. Townsend (2007): "Sources of TFP Growth," *Journal of Economic Theory*, 32(1), 179–221.

Klenow, P., and A. Rodriguez-Clare (1997): "The Neoclassical Revival in Growth Economics: Has it Gone Too Far?," in *NBER Macroeconomics Annual*, vol. 12, pp. 73–103.

Klette, T. J., and S. Kortum (2004): "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 112(5), 986–1018.

La Porta, R., and A. Shleifer (2009): "The Unofficial Economy and Economic Development," in *Brookings Papers on Economic Activity*, pp. 275–352.

Lagos, R. (2006): "A Model of TFP," *Review of Economic Studies*, 73(4), 983–1007.

Lentz, R., and D. T. Mortensen (2008): "An Empirical Model of Growth through Product Innovation," *Econometrica*, 76(6), 1317–1373.

Levinsohn, J., and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–341.

Lucas, R. E. (1978): "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 9(2), 508–523.

——— (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22(1), 3–42.

——— (1990): "Why Doesn't Capital Flow From Rich to Poor Countries," *American Economic Review*, 80(2), 92–96.

Luttmer, E. G. J. (2010): "Models of Growth and Firm Heterogeneity," *Annual Review of Economics*, 2, 547–576.

McKinsey-Global-Institute (2001): "India: The Growth Imperative," .

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

Melitz, M., and G. Ottaviano (2008): "Market Size, Trade and Productivity," *Review of Economic Studies*, 75(1), 295–316.

Midrigan, V., and D. Y. Xu (2010): "Finance and Misallocation: Evidence From Plant-Level Data," NBER Working Paper 15647.

Moll, B. (2010): "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?," Working Paper.

Parente, S. L., and E. C. Prescott (1994): "Barriers to Technology Adoption and Development," *Journal of Political Economy*, 102(2), 298–321.

Pavcnik, N. (2002): "Trade Liberalization, Exit and Productivity Improvements: Evidence from Chilean Plants," *Review of Economic Studies*, 69(1), 245–276.

Restuccia, D., and R. Rogerson (2008): "Policy Distortions and Aggregate Productivity with Heterogenous Establishments," *Review of Economic Dynamics*, 11(4), 707–720.

Sutton, J. (1997): "Gibrat's Legacy," *Journal of Economic Literature*, 35, 40–59.

5 Appendix

5.1 Data

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

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

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

To derive (20), simply note that

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

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

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

which is (22). Similarly,

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

so that

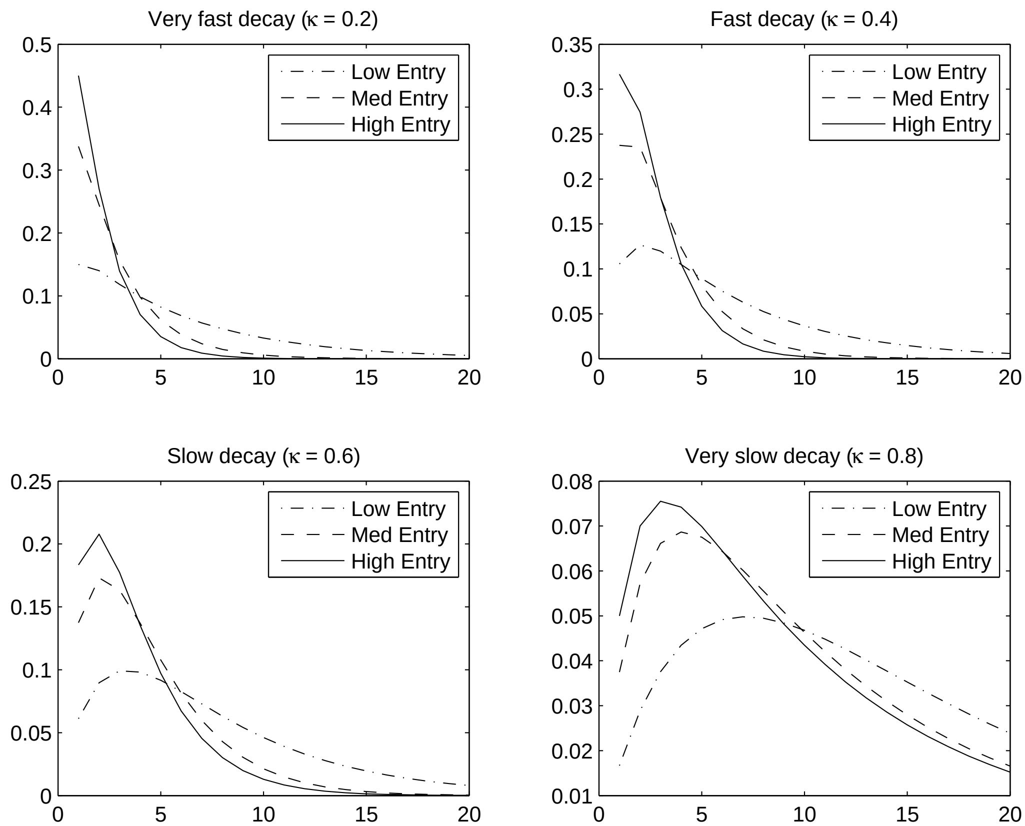
$$M(x) = \frac{\lambda^{-\int_0^1 \Delta(\nu)d\nu}}{\int_0^1 \lambda^{-\Delta(\nu)}d\nu} = \frac{1}{\Lambda(x)}\lambda^{-\int_0^1 \Delta(\nu)d\nu} = \lambda^{-\frac{1+x}{x}}\frac{\lambda-1+\lambda x}{x},$$

which is (23).

5.3 More general entry process: Derivation of (24)

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

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



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

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

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

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

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

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

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

5.4 Characterization of the BGP

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

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

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

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

the Bellman equation reads

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

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

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

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

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

so that upon substituting (72), we get that

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

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

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

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

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

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

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

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

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

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

To finally determine the equilibrium innovation level I and the entry intensity x, consider (31) and (32), i.e.

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

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

From (69),

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

as $t_x(x,I) < 0$ and $t_I(x,I) > 0$. From (70),

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

as

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

Hence, (31) and (32) define two continuous loci with different slopes. Additionally,

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

where $I^{FE}(x)$ is the locus implicitly defined by the free entry condition (31). Similarly,

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

where $I^{LM}(x)$ is locus implicitly defined by the labor market clearing condition (32). Hence, there is a unique intersection and therefore a unique equilibrium.

5.5 Deriving the Budget Constraint (34)

To derive (34), it is useful to think of households having access to two assets to transfer resources across time. They can either save in physical capital or they can save in a mutual fund, who owns the corporate sector of the economy and finances entrants by venture capital. This mutual fund pays an interest $r(t)$ for each dollar invested. Letting $A(t)$ be the savings in the mutual fund, the budget constraint of the representative household is given by

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

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

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

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

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

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

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

Substituting this into (71) yields

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

which is (34).

5.6 Static Misallocation and Dynamic Incentives

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

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

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

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

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

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

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

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

so that

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

The necessary and sufficient FOC for innovation choice is still

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

so that

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

Now suppose that $I=0$. Then

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

Along the steady state,

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

so that

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

which shows that

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

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

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

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

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

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

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

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

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

This is a parametric condition, which is satisfied if the fixed costs f are high, the productivity φ is low, the innovation step-size λ is low and the discount rate ρ is high, so that innovations, which pay off in the future, are valued little.

For this equilibrium to occur, we also require that there are no entry incentives. There is no entry if

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

Rearraning terms yields

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

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

5.7 A Simple Model of Credit Constraints

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

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

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

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

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

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

If $\mu = 0$, then

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

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

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

Hence, the firm is constrained if

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

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

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

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

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

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

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

To get (41), note that

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

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