MISALLOCATION: MARKUPS AND TECHNOLOGY

Abstract

The seminal paper by Hsieh and Klenow (2009) shows that aggregate TFP losses from misallocation are large, that misallocation is important in explaining international TFP differences, and that these losses can be quantified through factor productivity dispersions. Using micro data from Chile, Colombia, Indonesia, and Germany, we show that there is substantial correlation in factor productivities across factors and therefore propose to decompose dispersion in factor productivities in dispersion in technology and markup instead, which are orthogonal to each other. Relative to Germany, misallocation is larger in the developing economies and the TFP losses from misallocation are explained to 1/3 by larger technology and to 2/3 by larger markup dispersion. Finally, we discuss potential sources of markup and technology dispersion and how they can rationalize the observed dispersions as market outcomes.

JEL Classification: D24, E23, O47

Keywords: productivity, Development, Misallocation, Competition

Christian Bayer - christian.bayer@uni-bonn.de
Universität Bonn

Ariel Mecikovsky - amecikovsky@gmail.com
Ford Motor Company

Matthias Meier - m.meier@uni-mannheim.de
University of Mannheim

Acknowledgements
The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FTP/2007-2013) / ERC Grant agreement no. 282740. Ariel Mecikovsky and Matthias Meier thank the DFG for supporting their research as part of the research training group GRK 1707 "Heterogeneity, Risk and Economic Dynamics.'' Matthias Meier gratefully acknowledges support from the UniCredit & Universities Foundation, DAAD and DFG. Matthias Meier conducted parts of this paper while visiting NYU, and he would like to thank the department, especially Gianluca Violante, for its hospitality. Christian Bayer conducted part of the research while he was a Fernand Braudel Fellow at the EUI and would like to thank the EUI for its hospitality. We are grateful to Jörg Breitung, Simon Gilchrist, Matthias Kehrig, Dirk Krueger, Virgiliu Midrigan, Venky Venkateswaran for detailed comments and the participants at various conferences and seminars. We thank Deutsche Bundesbank, Instituto Nacional de Estadisticas de Chile} Badan Pusat Statistik, and Devesh Raval for providing us access to the firm\plant-level data from Germany, Chile, Indonesia, and Colombia, respectively.

Misallocation: Markups and Technology*

Christian Bayer,[†] Ariel M. Mecikovsky,[‡] Matthias Meier[§]

This version: October 20, 2018

Abstract

The seminal paper by Hsieh and Klenow (2009) shows that aggregate TFP losses from misallocation are large, that misallocation is important in explaining international TFP differences, and that these losses can be quantified through factor productivity dispersions. Using micro data from Chile, Colombia, Indonesia, and Germany, we show that there is substantial correlation in factor productivities across factors and therefore propose to decompose dispersion in factor productivities in dispersion in *technology* and *markup* instead, which are orthogonal to each other. Relative to Germany, misallocation is larger in the developing economies and the TFP losses from misallocation are explained to 1/3 by larger technology and to 2/3 by larger markup dispersion. Finally, we discuss potential sources of markup and technology dispersion and how they can rationalize the observed dispersions as market outcomes.

JEL-Codes: D24, E23, O47

Keywords: Productivity, Development, Misallocation, Competition.

*An earlier version of this paper was titled: "Dynamics of Factor Productivity Dispersion." The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FTP/2007-2013) / ERC Grant agreement no. 282740. Ariel Mecikovsky and Matthias Meier thank the DFG for supporting their research as part of the research training group GRK 1707 "Heterogeneity, Risk and Economic Dynamics." Matthias Meier gratefully acknowledges support from the UniCredit & Universities Foundation, DAAD and DFG. Matthias Meier conducted parts of this paper while visiting NYU, and he would like to thank the department, especially Gianluca Violante, for its hospitality. Christian Bayer conducted part of the research while he was a Fernand Braudel Fellow at the EUI and would like to thank the EUI for its hospitality. We are grateful to Jörg Breitung, Simon Gilchrist, Matthias Kehrig, Dirk Krueger, Virgiliu Midrigan, Venky Venkateswaran for detailed comments and the participants at various conferences and seminars. We thank *Deutsche Bundesbank*, *Instituto Nacional de Estadísticas de Chile*, *Badan Pusat Statistik*, and Devesh Raval for providing us access to the firm/plant-level data from Germany, Chile, Indonesia, and Colombia, respectively.

[†]CEPR, IZA, and Department of Economics, Universität Bonn. Address: Adenauerallee 24-42, 53113 Bonn, Germany. Email: *christian.bayer@uni-bonn.de.*

[‡]Ford Motor Company.

[§]Department of Economics, Universität Mannheim. Address: Block L7, 3-5, 68161 Mannheim, Germany. Email: *m.meier@uni-mannheim.de.*

1 Introduction

The allocation of factors to their most productive use is a key determinant of economic prosperity (Jones, 2016). First-best efficiency requires that factors produce the same marginal revenue across all production units. Therefore, it has become a popular approach to quantify the aggregate productivity losses from misallocation by estimating in micro data deviations (wedges) from the first-order conditions that characterize the efficient factor demand. Following the seminal work by Hsieh and Klenow (2009), many studies show these distortions to be substantial finding large dispersions of factor productivity even within industries.[1]

This paper suggests to look at misallocation from a somewhat different angle. We ask whether a plant is inefficient in *how it produces* a given level of output (i.e., using the wrong factor mix, in short technology), and whether it is inefficient in *how much it produces* (such that it has deviations in markups from the optimal level). In other words, we rotate wedges in the first-order conditions into differences/wedges in markups and technologies. Whereas wedges in the first-order conditions are correlated across factors, which makes it somewhat hard to interpret them, we show that our rotation yields uncorrelated wedges.[2] In addition we provide evidence on the persistence of measured misallocation at the plant level. Finally, we discuss a variety of potential explanations for misallocation.

We use micro data from four countries—Germany (firm-level), Chile, Colombia, and Indonesia (plant-level)—and obtain five results: First and foremost, misallocation is very persistent at the plant level. Between 66% and 90% of all identified productivity losses come from persistent misallocation within 4 digit industries. Second, all considered dispersion measures are substantially larger in the three developing economies vis-à-vis Germany. Third, single-factor (capital and labor) productivity wedges are correlated among each other, while wedges in "technology" and "markup" are close to orthogonal. Fourth, dispersion of factor productivities is mainly the result of persistent technology differences, and not markup dispersion. Yet, markup dispersion is much more costly in terms of measured productivity losses than is dispersion in technology. Overall, we estimate that aggregate TFP could be between 12% and 44% higher under the first-best efficient allocation of production factors, i.e., absent *any* misallocation. Two-third of the measured efficiency loss through misallocation comes from firms choosing a suboptimal

[1] See Restuccia and Rogerson (2008), Hsieh and Klenow (2009), Peters (2016), Asker et al. (2014), Gopinath et al. (2017), and Restuccia and Santaeulalia-Llopis (2017) to name a few.

[2] Our approach can be viewed as a decomposition of the wedges Hsieh and Klenow (2009) identify without changing the identification of misallocation itself.

scale of operations compared to the industry benchmark markup. Fifth, we find that persistent technology differences are significantly less dispersed in more competitive environments. In other words, firms with less competitive pressure can maintain inefficient production technologies more easily then firms with strong competition.

This empirical evidence is informative about potential explanations of measured misallocation. One popular view is that misallocation results from policy-induced distortions. We instead ask whether there may be market-based explanations, too. First, persistent differences in markups can be the result of imperfect competition that is not monopolistic. Limit pricing is one such example. Second, we argue that less stable demand in the developing countries may be the source of more volatile markups. Third, we argue that adjustment costs on factor mixes, adjustments costs on technology, are likely the key ingredient to explain dispersions in "technology". Consistent with this explanation, we show that these dispersions are larger in less competitive environments.

Let us be more specific about the potential reasons of persistent markup differences. For example, the coexistence of niche and mass products in narrowly defined industries can lead to markup dispersion (e.g., Bar-Isaac et al., 2012) – think of generics vs. patented pharmaceuticals.[3] Similarly, persistent markup differences may arise due to slow customer base accumulation and staggered entry in growing industries (see, e.g., Gourio and Rudanko, 2014). A particularly simple setup to think about persistent markup dispersions is limit pricing as in Peters (2016). Importantly, in such context markup dispersion may be the result of increased competition or more innovation when either competition or innovation is scarce. As such, greater markup dispersion may be a phenomenon of faster growth, and not its impediment.

Similarly, there are good established theories for transitory markup differences. Rigid prices naturally lead to markup differences if costs evolve stochastically. In monetary economics, price rigidities are a channel through which inflation and inflation volatility leads to welfare losses, see Calvo (1983) or Sheshinski and Weiss (1977). In fact, higher inflation and inflation volatility might be behind our finding of higher transitory markup dispersions in developing countries. Another potential explanation of cross-country differences is that demand at the firm/plant level is less stable in these countries (see Asker et al., 2014).

Lastly, our finding of persistent technology dispersion across plants links to the traditional putty-clay theory (Johansen, 1959), which has been advocated to address a

[3]Our results are complementary to Haltiwanger et al. (2018), who study departures from CES demand and constant marginal costs in the Hsieh and Klenow (2009) framework. A related point regarding the relationship between market structure and misallocation has also been raised by Asker et al. (2017) studying the role of OPEC on the oil market.

broad array of other empirical phenomena (Gilchrist and Williams, 2000, 2005; Gourio, 2011). Technology differences might, for example, be the result of putty-clay technology choice under factor price uncertainty and lump-sum technology adjustment costs as in Kaboski (2005).[4] The fact that firms under less competitive pressure have more disperse technology speaks in this direction. These firms can roll over to their customers some of their higher costs of production from an inefficient technology. This aligns also well with the fact that technology dispersion decreases in plant size, if technology adoption has increasing returns to scale or if credit constraints affect technology adoption (see, e.g., Banerjee and Duflo, 2005; Midrigan and Xu, 2014). Cross-country differences in technology dispersion may also reflect the fact that relative factor prices are more variable in developing countries, whether because of aggregate variability, because of local market conditions, or because of financial frictions binding for some firms. Relatedly, David et al. (2018) show that firms persistently differ in their exposure to aggregate conditions, which translates into differential costs of capital, and thus technology differences.

If the dispersions in technology are indeed an indicator of substantial frictions that firms are facing in adjusting their production technology, then this has an important implication for empirical estimates of substitution elasticities. Compared to the frictionless benchmark, firms will mute their technology adjustment in response to relative factor-price changes if these are mean reverting. Consequently, simple regressions of capital intensities on relative factor prices will fail to identify the long-run elasticity of substitution even if they manage to identify exogenous price changes (see, e.g., Raval (2014) or Oberfield and Raval (2014) for recent contributions or Chirinko (2008) for an overview) and hence the estimated substitution elasticities are subject to a downward bias. In fact, we provide evidence that this downward bias is likely substantial and suggest an IV strategy to identify the long-run elasticity. When we regress factor intensities on persistent factor price changes, the estimated substitution elasticity roughly triples.[5] This not only has important implications for income shares (see, e.g., Solow, 1956; Piketty, 2011, 2014; Karabarbounis and Neiman, 2013) but is also key for the efficiency losses from technology dispersion. The more substitutable factors are the smaller are losses from a suboptimal capital-labor ratio.

The remainder of this paper is organized as follows: Section 2 provides empirical results, Section 3 decomposes the observed dispersions in TFP losses, and Section 4 revisits the role of market structure. Section 5 concludes and an appendix follows.

[4]A related explanation of technology dispersion is in Uras and Wang (2017), who consider a model in which firms make distorted choices of their CES production technology weights on capital and labor.

[5]Chirinko and Mallick (2017) provide an alternative estimation strategy to identify the long-run elasticity by filtering out transitory variation.

2 Dissecting factor productivity dispersions

2.1 Data description

We study dispersion in factor revenue productivity and its constituting components using firm-level data from Germany and plant-level data from Chile, Colombia and Indonesia. For Germany, we use the balance-sheet database of the Bundesbank, USTAN, which is private-sector, annual firm-level data available for 26 years (1973-1998). For Chile, Colombia and Indonesia, we have plant-level data from the ENIA survey for 1995-2007, the EAM census for 1977-1991 and the IBS dataset for 1988-2010, respectively. All data sets are focused on the manufacturing sector, with the exception of Germany, which provides information for the entire private non-financial business sector. More details on the data are provided in Appendix A.1.

In preparing the data for our analysis, we treat the various data sets in the most comparable way. From each database, we use a firm's/plant's wage bill, value-added, capital stock in book or current value, and its four-digit industry code. To obtain economically consistent capital series for each firm/plant, we re-calculate capital stocks using the perpetual inventory method whenever capital stocks are reported in book values. We exploit information on capital disaggregated into structures and equipment, which allows us to control for heterogeneity in capital composition across firms/plants.

We further need information on the depreciation and real interest rates. We do not rely on depreciation as reported by firms/plants, as it is potentially biased for tax purposes. Instead, we use economic depreciation rates by type of capital good obtained from National Statistics.[6] We then estimate a firm/plant-year specific depreciation rate, δ_{it}, that capital mix of structures and equipment. We set the real rate to 5% for all economies. This implies user costs of capital $R_{it} = 5\% + \delta_{it}$. In generating cross-sectional statistics, we control for time variations in user costs by taking out four-digit industry-year fixed effects. The data treatment and sample selection are described in detail in Appendix A.2.

2.2 Persistent and transitory factor productivities

We compute average factor productivities per firm/plant (i) and year (t) using the reported value added at current prices, $p_{it}y_{it}$; labor expenses, $W_{it}L_{it}$ as reported in the

[6] Depreciation rates for equipment and structures are obtained from *Volkswirtschaftliche Gesamtrechnung* (VGR) for Germany, and from Henriquez (2008) for Chile. We use the Chilean data for Colombia and Indonesia in lieu of national data, which are not available. The depreciation rates are 15.1% (equipment) and 3.3% (structures) in Germany, and 10.5% (equipment) and 4.4% (structures) for Chile.

profit and loss statements; and capital expenses, $R_{it}K_{it}$, computed following the steps detailed above. Taking logs, we define revenue productivities of labor and capital:

$$\alpha_{it}^{L} := \log(p_{it}y_{it}) - \log(W_{it}L_{it}); \qquad \alpha_{it}^{K} := \log(p_{it}y_{it}) - \log(R_{it}K_{it}). \tag{1}$$

Using expenditures and value added implicitly controls for quality differences in both inputs and outputs (c.f. Hsieh and Klenow, 2009). We remove four-digit industry-year fixed effects from the data.

For any of these variables, say, x_{it}, we distinguish between persistent and transitory deviations from the industry-year mean. We identify the persistent component $\bar{x}_{it}$ as 5-year moving averages, and the transitory component $\hat{x}_{it}$ as deviations thereof:

$$\bar{x}_{it} := \frac{1}{5}\sum_{s=-2}^{2} x_{it+s}, \quad \hat{x}_{it} := x_{it} - \bar{x}_{it}. \tag{2}$$

We further consider a 9-year moving-average filter, set up analogously to equation 2, to identify components of particularly high persistence.

The first panel of Table 1 reports standard deviations and correlation for labor and capital productivity in all four countries. Three observations stand out: First, capital and labor productivity are positively correlated in the transitory component ($\rho \approx 50\%$), while they tend to be negatively correlated in the persistent component ($\rho \approx -10\%$). Second, the persistent components in productivity explain the vast majority of cross-sectional productivity differences (between 60% and 92% for labor and between 79% and 94% for capital). Third, the developing economies show larger productivity dispersions. The differences in the transitory component are more pronounced compared to the persistent component.[7]

To shed further light on how long-lived factor productivity differences between firms/plants are, we use a nine-year moving-average filter. The findings are documented in Table 9 in the Appendix. The striking observation is that even at such long horizon, the persistent component still accounts for at least half of the total variance in labor productivity and at least two third of the total variance in capital productivity. In other words, a large fraction of the observed deviations in firm/plant-specific factor productivities from their industry-year specific mean appear not to revert back to mean but rather reflect permanent differences.

Finally, the finding of a positive correlation between labor and capital productivity

[7] These findings are robust, both to weighting firm/plant-level observation by value added, see Table 8, and to replacing the five-year moving-average filter by the HP filter with $\lambda = 6.25$, see Table 7.

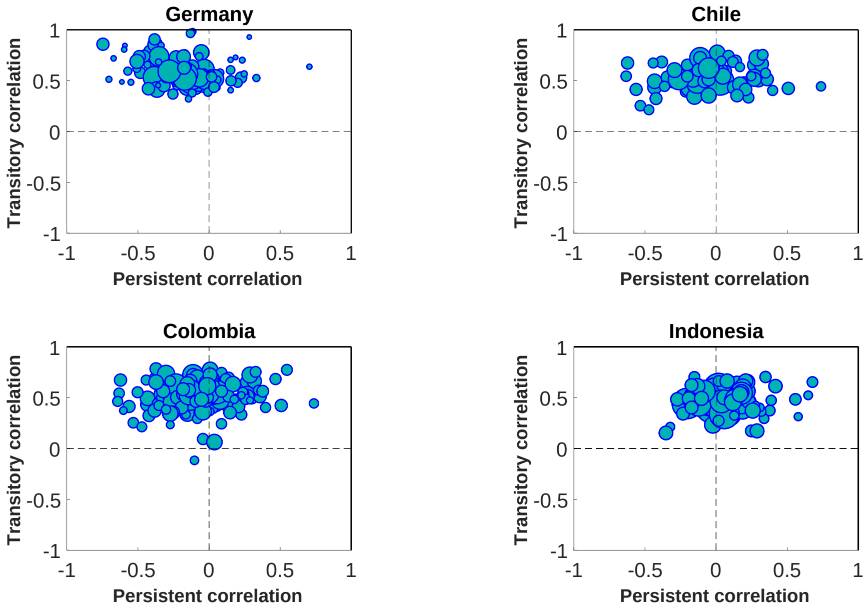
Table 1: Transitory and persistent components of factor productivities, markup and technology

	$\text{std}(\hat{\alpha}^L_{it})$	$\text{std}(\hat{\alpha}^K_{it})$	$\rho(\hat{\alpha}^L_{it}, \hat{\alpha}^K_{it})$	$\text{std}(\bar{\alpha}^L_{it})$	$\text{std}(\bar{\alpha}^K_{it})$	$\rho(\bar{\alpha}^L_{it}, \bar{\alpha}^K_{it})$
	Transitory Component			Persistent Component		
DE	0.066	0.119	0.352	0.229	0.456	-0.207
	(0.000)	(0.001)	(0.002)	(0.002)	(0.004)	(0.004)
CL	0.226	0.310	0.545	0.294	0.613	-0.076
	(0.006)	(0.007)	(0.014)	(0.011)	(0.024)	(0.020)
CO	0.153	0.182	0.552	0.273	0.681	-0.074
	(0.003)	(0.003)	(0.010)	(0.007)	(0.021)	(0.016)
ID	0.288	0.415	0.483	0.353	0.736	0.014
	(0.004)	(0.004)	(0.006)	(0.005)	(0.011)	(0.009)

	$\text{std}(\hat{\mu}_{it})$	$\text{std}(\hat{\kappa}_{it})$	$\rho(\hat{\mu}_{it}, \hat{\kappa}_{it})$	$\text{std}(\bar{\mu}_{it})$	$\text{std}(\bar{\kappa}_{it})$	$\rho(\bar{\mu}_{it}, \bar{\kappa}_{it})$
	Transitory Component			Persistent Component		
DE	0.064	0.114	-0.155	0.172	0.551	0.062
	(0.000)	(0.001)	(0.002)	(0.001)	(0.004)	(0.004)
CL	0.219	0.266	-0.082	0.241	0.700	-0.084
	(0.005)	(0.008)	(0.014)	(0.007)	(0.027)	(0.019)
CO	0.145	0.160	-0.085	0.287	0.752	-0.507
	(0.003)	(0.004)	(0.011)	(0.006)	(0.022)	(0.013)
ID	0.278	0.373	-0.097	0.310	0.812	-0.102
	(0.003)	(0.004)	(0.006)	(0.004)	(0.012)	(0.008)

Notes: Cross-sectional standard deviations (std) and correlation (ρ) of transitory and persistent components of labor and capital productivity, α^L_{it} and α^K_{it} as in (1), capital intensity, κ_{it}, and markup, μ_{it}, as in (3) and (4). DE: Germany, CL: Chile, CO: Colombia, ID: Indonesia. Transitory and persistent components are obtained by applying a five-year moving average filter. Factor productivities are demeaned by 4-digit industry and year, and expressed in logs. In parentheses: Clustered standard errors at the firm/plant level.

Figure 1: Correlations of factor productivities by four-digit industry



Notes: *Transitory (Persistent) Correlation*: Correlation between the transitory (persistent) component of labor and capital productivity at the firm/plant level, controlling for year fixed effects. Each circle represents a four-digit industry, where the size of a circle reflects aggregate employment in that industry. For this figure, we restrict industries to those that include at least 20 firms/plants.

in the short run and a negative/zero correlation in the long run does not only hold within four-digit industries. In fact, we observe the same correlation pattern for the bulk of four-digit industries, see Figure 1.

2.3 Rotating the productivity data: markups and technology

One alternative way to look at the data is to represent deviations in factor productivities by deviations in markups and technology. This is similar to looking at the average of capital and labor productivity and their difference. Our markup measure is value added relative to total expenditures on labor and capital

$$\mu_{it} := \log(p_{it}y_{it}) - \log(R_{it}K_{it} + W_{it}L_{it}) \tag{3}$$

and our technology measure is the factor price-weighted capital intensity

$$\kappa_{it} := \log(R_{it}K_{it}) - \log(W_{it}L_{it}). \tag{4}$$

Using a CES production function, labor and capital productivity can be mapped into markup and technology and vice-versa.

Similar to factor productivities, we do not view our markup and technology measures as structural parameters. We also do not view these as wedges in the spirit of Hsieh and Klenow (2009). Instead markup and technology are *labels* attached to the reduced-form objects constructed through equations (3) and (4). However, measured markups and technology differences lend themselves to economic interpretation. They are instructive about the potential frictions behind them.

The second panel of Table 1 documents transitory and persistent markup and capital intensity differences for all countries. The first observation is that capital intensities and markups are virtually orthogonal (with persistent markups and capital intensities in Colombia being the single exception). This suggests that analyzing the misallocation data in terms of markups and capital intensities is the right perspective.

Second, differences in capital intensity are very persistent. The transitory component makes up only between 4% (Germany) and 17% (Indonesia) of the total variance of capital intensities. At the same time, persistent differences in capital intensity are substantially more dispersed in Chile, Colombia, and Indonesia than they are in Germany with variances being twice as high in Indonesia compared to Germany.

On the contrary, the dispersion of persistent cross-sectional markup differences is strikingly similar across countries, and transitory differences in markups are an important component of the total cross-sectional variance of markups – at least in the developing economies (30% in Colombia, 50% in Chile and Indonesia) but less so in Germany (12%). This may be related to demand being less stable in the developing economies. In fact, the cross-sectional standard deviation of value-added growth (not reported) is two to four times larger in these economies than in Germany.

These empirical findings are again highly robust, both to weighting and the HP filter, see Tables 8 and 7. Further, especially for technology a large share of deviations from industry-year mean are very long-lived. In fact, most of the deviations persist well beyond five year as the results from the nine-year moving-average filter in Table 9 show. This suggests that a large share of the technology differences rather reflect permanent differences across firms/plants within industries.

In the appendix, we also investigate the effect our data treatment has on the measured technology and markup dispersions. We report dispersions conditional on various data treatment steps; see Tables 11, 12, and 13 in the appendix. The estimated TFP losses can be up to four times larger when we drop various data-cleaning steps.

3 Implications for aggregate productivity

3.1 TFP loss decomposition

While the magnitudes of dispersion in technology and markup within narrowly-defined industries are interesting in their own right, they are hard to interpret in terms of their macroeconomic importance. To obtain such an interpretation, we view them through the lense of a model that is motivated by and similar to the model in Hsieh and Klenow (2009). The important difference in our approach is that we do not assume certain wedges to account for the observed dispersion. Instead, we directly map deviations of technology and markup from their optimal values into aggregate total factor productivity (TFP). In doing so, we use a second order approximation to the unit-costs of production that makes use of the fact that markups and technology are approximately orthogonal in the data.

We consider a unit mass of ex-ante identical firms, indexed by i. Each firm operates a constant elasticity of substitution (CES) production function

$$y_i = f(k_i)L_i \ , \quad f(k_i) = [\alpha k_i^{\frac{\rho-1}{\rho}} + (1-\alpha)]^{\frac{\rho}{\rho-1}}, \qquad \text{(Assumption 1)}$$

where ρ is the elasticity of substitution between labor L_i and capital K_i, and $k_i = K_i/L_i$. We assume households have Dixit-Stiglitz preferences over output varieties y_i, such that aggregate output is

$$Y = \left[\int (z_i y_i)^{\frac{\eta-1}{\eta}} di\right]^{\frac{\eta}{\eta-1}}, \qquad \text{(Assumption 2)}$$

where z_i denotes a demand shifter, and η is the elasticity of substitution between varieties. Firms engage in monopolistic competition, such that firm i's demand curve is given by

$$\begin{aligned} y_i &= z_i^{\eta-1} (p_i/P)^{-\eta} Y, \\ \text{where } P &= \left(\int z_i^{\eta-1} p_i^{1-\eta} di\right)^{\frac{1}{1-\eta}} \end{aligned} \qquad \text{(Assumption 3)}$$

such that the price of any single firm i has an effect only on that firm's demand.

All firms face the same wage, W, and user cost of capital, R. This implies that their marginal costs of production are a function of their capital-labor ratio, k_i,

$$c_i = c(k_i) = \frac{W + Rk_i}{f(k_i)}. \tag{5}$$

The (gross) markup m_i is defined as

$$m_i = p_i/c_i. \tag{6}$$

This setup makes transparent that firms and consumers share the benefits of lower production costs. In turn, the lower the elasticity of substitution η the more of a cost increase the firm can offload onto the consumer and the higher the ideal markup of that firm would be. We have not specified the trade-offs firms face in setting prices and technology, but the model can explain that firms in less competitive (higher markup) markets have substantially larger technology dispersions.

We use this model setup to derive the TFP losses that of dispersion in technology and markups. Our benchmark is an economy in which all firms use the unit-cost minimizing capital intensity $k^* = \arg\min_k c(k)$, charge unit markup $m = 1$, and face a productivity/demand shifter of $z = 1$. We compute TFP losses using a second-order approximation around the benchmark and with respect to $\kappa = \log(k)$, $\zeta = \log(z)$, and $\mu = \log(m)$. This yields

$$\text{TFP loss} \approx \mathbb{E}(\zeta) + (\eta - 1)\mathbb{V}(\zeta) - \eta\mathbb{V}(\mu) - \frac{1}{2\rho}s^*(1 - s^*)\mathbb{V}(\kappa), \tag{7}$$

where $\mathbb{E}$ and $\mathbb{V}$ denote the cross-sectional expectations and variance operator, respectively. s^* is the capital expenditure share in the cost-minimizing optimum

$$s^* = Rk^*/(W + Rk^*).$$

The first term, $\mathbb{E}(\zeta)$, in (7) is aggregate productivity in the absence of any heterogeneity across production units; the second term, $(\eta - 1)\mathbb{V}(\zeta)$, reflects the Oi-Hartmann-Abel effect of productivity dispersion. If individual productivities ζ are more dispersed, aggregate productivity goes up/average costs fall. The third term, $-\eta\mathbb{V}(\mu)$, captures the loss from markup dispersions, from deviations in allocative efficiency. The larger substitution elasticity η between varieties, the larger this loss.[8] The last term, $-\frac{1}{2\rho}s^*(1 - s^*)\mathbb{V}(\kappa)$,

[8]Note that a larger η raises the quantity dispersion for some given price dispersion that lowers productivity. At the same time, a larger η lowers the productivity losses for some given quantity dispersion because different varieties become more substitutable. It is a feature of the CES aggregate that the

captures how the dispersion of capital intensity lowers TFP. The higher the substitution elasticity ρ, the smaller is the loss from using an inefficient technology. A more detailed derivation is provided in Appendix A.7. Broadly speaking, our approximation makes clear that an optimal allocation answers two separate questions: $\mathbb{V}(\kappa)$ measures misallocation in the sense of *how things are produced* and $\mathbb{V}(\mu)$ measures misallocation in the sense of *what is produced*.

3.2 Parameterization

To map the dispersions in the data into aggregate productivity losses, we need to paramterize s^*, ρ, and η. Assuming s^* is well approximated by the average cost share across producers and time, we estimate s^* directly from the micro data. We obtain a capital share of 21% (Germany), 40% (Colombia), 32% (Chile), and 23% (Indonesia).

We estimate the elasticity of substitution between labor and capital, ρ, from time-series information on the aggregate capital intensity and the relative factor price. In a frictionless economic environment, the elasticity is determined by the contemporaneous correlation between these variables. However, the identification is problematic in the presence of frictions that prevent the immediate adjustment of production factors: The contemporaneous response of the capital intensity to price movements (short-run elasticity) then differs from the long-run elasticity, which we estimate from aggregate cross country data, see Section 4.3. We obtain an estimated elasticity of 1.28. Our estimate is well within the range of estimated values in the literature from 0.4 (Chirinko and Mallick, 2017) to 3.4 (Ramirez Verdugo, 2005).

Finally, we follow Hsieh and Klenow (2009) in setting the remaining elasticity parameter to $\eta = 3$. This choice is conservative and restricts the role of misallocation in explaining cross-country productivity differences. Recall that higher elasticities of substitution increase the productivity losses from markup dispersion.

3.3 Results

Table 2 shows the loss in aggregate productivity due to dispersion in markups and technology. Through the lense of our model, a direct consequence of larger dispersions in developing economies is lower TFP. Quantitatively, the implied TFP differences are of substantial importance. For example, misallocation in markups and technology accounts for a 32% difference in aggregate TFP between Germany and Indonesia.

former effect dominates.

Table 2: TFP losses (in %)

DE	CL	CO	ID
12.1	28.4	28.6	43.8

Notes: TFP losses are computed using equation (7).

Next we decompose the total loss in TFP into losses from transitory and persistent variation in markups and technology, respectively; see the first panel of Table 3. Persistent technology differences, transitory markup differences and persistent markup differences are each important, while transitory technology differences are negligible for productivity lost through misallocation. There are substantial differences across countries in the importance of the three other components. In Germany more than half of the misallocation losses come from technology dispersion. In Indonesia, it is only 1/3, despite the far larger dispersion in technologies. There, as in Chile, short-run markup fluctuations are important and make up 1/4 of the productivity losses. Persistent markup differences across production units make up a third of all efficiency losses across all countries.

The second panel of Table 3 puts the productivity losses of the developing economies in a perspective relative to the German economy. On average transitory markup dispersion, persistent markup dispersion, and persistent technology dispersion each account for 1/3 of the TFP loss vis-a-vis Germany. Still, there are differences across the three developing countries. For Chile, the largest contributor in relative terms is the transitory markup dispersion; for Colombia, it is the persistent markup dispersion; and for Indonesia, both transitory and persistent markup dispersions are equally important.

4 Explanations of markup and technology dispersion

Within countries, the strongest forces of misallocation in terms of TFP losses are persistent markup differences and persistent differences in the factor mix—technology differences—, across firms in narrowly-defined industries. In the developing countries, another important factor of productivity losses are short-lasting differences in markups. Next, we ask to what extent established theories of imperfect competition, price setting and real frictions can account for these dispersions. In the context of these theories, we further ask whether misallocation, being an impediment to growth, can be addressed by

Table 3: Decomposition of TFP losses

	Transitory Component $(\hat{\mu}_{it})$	$(\hat{\kappa}_{it})$	Persistent Component $(\bar{\mu}_{it})$	$(\bar{\kappa}_{it})$
	Within-country decomposition of TFP losses (in %)			
DE	5.1	2.4	36.7	55.8
CL	25.4	5.5	30.7	38.4
CO	11.0	2.0	43.1	43.9
ID	26.5	7.1	32.9	33.5
	Accounting for the difference with Germany (in %)			
CL	40.4	7.9	26.3	25.4
CO	15.3	1.7	47.8	35.2
ID	34.7	8.9	31.5	25.0

Notes: Capital intensities, κ_{it}, and markups, μ_{it}, as defined in (3) and (4). See notes of Table 1 for further explanation.

simple policies.

4.1 Persistent markup differences as a result of market structure

Differences of markups across firms/plants are not all transitory. Instead, persistent markup dispersion is comparable in margnitude to transitory markup dispersion. Together they create about 70% of the relative TFP losses in developing economies. Thus, markup differences between firms/plants (within narrowly-defined industries) are key to understanding misallocation. We propose a joint explanation of transitory and persistent markup dispersions through differences in the competitive environment firms face and through differences in innovation.

Suppose we start from the setup in Section 3.1, in which every firm is the monopolist supplier of a variety. There are further no frictions such that all firms charge their optimal, identical markup. Now suppose in some variety sectors there is entry of similarly productive firms. As a result of increased competition, markups in those sectors will fall, and the cross-sectional dispersion of markups rises. Persistent markup differences may thus be the result of market structure. If firms (persistently) face different degrees of competition, this generates markup dispersion. Note however, that in such model economy, more markup dispersion may reflect more competition, less market power, and thus higher welfare. This stands in stark contrast to the conclusions drawn in Section 3.1.

Now suppose a firm that faces a competitor within its variety sector innovates. If the firm's productivity grows relative to its competitor, this will raise markups. If few firms innovate, then more innovation lowers markup dispersion. Similar to entry, innovation may explain a positive correlation between aggregate productivity growth and markup dispersion.In Appendix A.8, we formalize the mechanisms discussed here through a stylized model. We also carefully discuss the implications for aggregate productivity when introducing entry within variety markets and innovation in the model of Section 3.1.

To the extent that firms in developing economies are further away from the technology frontier, we may expect more frequent introduction of new technologies, either by incumbents or entering firms. According to the view just sketched, the need for firm-entry by itself may explain larger markup dispersion. In this case, an increase in markup dispersion can well be a byproduct of fast growth through increased entry and declining average markups through more competition.

4.2 Transitory markup differences as a result of rigid prices, inflation, and unstable demand

While persistent markup differences require us to think about market structure, transitory differences in markup are more likely the result of nominal or real rigidities. In fact, if prices are adjusted infrequently, then shocks to firm-specific profitability change the markup. Similarly, price setting frictions à la Calvo (1983) or Sheshinski and Weiss (1977) give rise to transitory markup differences across firms even in response to common aggregate shocks. Using the formula for price dispersion in Wooford (2011) yields a markup dispersion of roughly $\frac{\sqrt{\theta}}{1-\theta}\pi$, where $1-\theta$ is the probability to adjust prices and π is the inflation rate per period. The average annual inflation rates are 3%, 6%, 23% and 9% for Germany, Chile, Colombia, and Indonesia, respectively computed over the periods we have data for. If prices are adjusted at equal frequency across countries, different average inflation rates can account for some of the observed markup differences.[9]

Yet, there are other well known factors that can add to the dispersion of markups, like

[9] At quarterly frequency, suppose $\theta = 1/4$ so that the half life of price adjustments is four quarters. After computing quarterly inflation rates, we obtain markup dispersion of 0.03, 0.05, 0.18 and 0.08 for Germany, Chile, Colombia, and Indonesia.

idiosyncratic profitability shocks (demand and productivity). In fact, while the dispersion of idiosyncratic profitability shocks is estimated to be 0.095 in Germany (Bachmann and Bayer, 2013), an estimate for Colombia by Eslava et al. (2013) implies a shock dispersion of 0.24.[10] If it takes time for firms to adjust their prices and/or quantities, these differences in the stability of profitability must translate into larger markup dispersion. Quantitatively, the differences in demand stability translate into sizable differences in markup dispersion.[11] Examples of such frictions are time-to-build, time-to-hire, or costly customer accumulation, see, e.g., Gourio and Rudanko (2014).

Given the range of frictions involved and the strength of a simple price friction, it is if anything surprising that the transitory markup dispersions are not much higher in the developing economies. Of course, it remains an open question, why profitability is so much less stable in developing economies.

4.3 Technology adjustment friction

While the markup dispersions reflect differences in *how much plants produce* relative to their marginal cost curve, dispersion in the capital labor ratio reflect differences in *how plants produce*, in how they combine capital and labor. It has been a longstanding view that production technology is less moldable in the short run than in the long run. In turn, this means that firms need to pay costs to adjust their capital intensity of production. These costs may reflect the need to train workers and remodel the production process when more capital is used. Such a putty-clay model of production (Johansen, 1959) can lead to dispersions in production technology when adjustment is not perfectly synchronized across plants. In line with this view, Merz and Yashiv (2007) estimate that it is substantially cheaper for a firm to adjust both capital and labor simultaneously than doing this sequentially. This view of a rigidity in changing the capital-labor mix at the plant level has important further implications.

First, it implies that idiosyncratic differences in factor prices can lead to long-lasting differences in production technology. For example, a capital-intense production technol-

[10] In similar vein, we have estimated the dispersion in TFP growth across countries. In particular, we construct Solow residuals. Taking out four-digit industry and year fixed effects, the standard deviations of first differences in Solow residuals are 0.13, 0.40, 0.34, 0.49 for Germany, Chile, Colombia, and Indonesia. The order of standard deviations perfectly matches the order of estimated markup dispersions.

[11] Suppose firms face a downward-sloping (CES) demand curve, $p = \frac{1}{1-\xi} z^{\xi} y^{-\xi}$, where z is stochastic profitability. Assume firms face constant unit costs $c = 1$. If firms choose how much to produce, y, before observing shocks to z, the profit maximizing policy is $y = \mathbb{E}[z^{\xi}]^{1/\xi}$. Realized markups are $\frac{py}{cy} = \frac{1}{1-\xi} \frac{z^{\xi}}{\mathbb{E}[z^{\xi}]}$ and the standard deviation of log markups is $\xi/(1-\xi)\sigma_z$, where σ_z is the dispersion of innovations to z. If we set $\xi = 1/4$ (average gross markup of 1.33), the profitability shock dispersions for Germany and Colombia imply markup dispersions of 0.03 and 0.08, respectively.

Table 4: Persistent technology dispersion by firm/plant characteristics

	Markups		Size		Age	
	Bottom Quintile	Top Quintile	Bottom Quintile	Top Quintile	Young	Old
DE	0.545	0.622	0.610	0.509	n.a.	n.a.
	(0.010)	(0.010)	(0.009)	(0.011)		
CL	0.646	0.743	0.808	0.658	n.a.	n.a.
	(0.052)	(0.063)	(0.070)	(0.050)		
CO	0.505	0.689	0.815	0.736	0.767	0.771
	(0.028)	(0.053)	(0.049)	(0.050)	(0.092)	(0.046)
ID	0.763	0.831	0.855	0.847	0.794	0.839
	(0.027)	(0.031)	(0.029)	(0.029)	(0.046)	(0.032)

Notes: Bottom (top) markup quintile: firm/plant average markup below the 20th percentile (above the 80th percentile). Old (young): Plant age below 4 years (above 15 years). Bottom (top) size quintile: firm/plant average employment below the 20th percentile (above 80th percentile). The micro data from Germany and Chile do not include age. See notes of Table 1 for further explanation.

ogy is more expensive for a financially constrained firm than for an unconstrained one. The constrained firm would not adopt an otherwise more efficient technology (c.f. Midrigan and Xu, 2014). Technology adoption costs in combination with financial frictions may explain persistent technology dispersion.

Second, if technology dispersion comes from active adjustment choices, then not all firms have the same incentive to adjust their technology. In particular, plants with more market power can offload a larger share of the costs associated with operating a suboptimal technology onto their consumers. Hence, these plants will tolerate larger deviations from cost-minimizing technologies.

To test these first two implication, we split the sample according to firm/plant characteristics – age, size, and, importantly, a firm's average markup – and compute again the dispersions of the persistent component of capital intensity; see Table 4. What stands out is the sample split along the average firm/plant markup. The highest markup quintile exhibits between 30% and 60% higher capital intensity dispersions (in terms of variances) than the lowest markup quintile. Yet, the capital-labor ratios tend to be more dispersed also among smaller plants, a group where a larger share can be expected to be financially constrained. These results remain significant also when simultaneously controlling for age, size, and market power, see Table 10 in the Appendix.

A third implication of technology adjustment costs is that simple OLS regressions of capital intensity on relative factor prices will not identify the true technological long-run elasticity of substitution but rather a short-run elasticity that combines adjustment costs, potential effects of mean reversion in relative factor prices and the true technological elasticity. First, plants will react sluggishly to changes in relative factor prices if there are adjustment costs. Second and more importantly, forward-looking plants react stronger to changes in factor prices which can be expected to be permanent or persistent than to transitory changes.

We test this implication of technology adjustment costs using aggregate data from the Penn World Table 8.0, see Feenstra et al. (2015). We use data from 99 countries over the years 1956-2002. We regress the aggregate capital to labor ratio (in logs) on the log wage to interest rate ratio. For total labor we use total hours worked. If hours worked information is missing, we impute hours worked using the number of employed and the average hours worked per employed in those countries with available data. We run four regressions: First, a simple OLS regression, second a regression with country fixed effects (within country), third a regression of country average capital intensities on average wages (between country regression), and fourth and finally a regression, where we instrument relative factor price changes by the country's top marginal income tax on domestic corporations. Again we control for country fixed effects.[12]

Table 5 summarizes the results of this estimation exercise. The OLS estimate of the elasticity of substitution is 0.68, well below one. The within-country estimate is even smaller at 0.43. However, this estimator identifies the elasticity from deviations from the global trend and the country average. By construction this takes out most long-lived differences and increases the importance of short lived fluctuations in relative factor prices for the regression. A little unsurprisingly, when we regress capital intensities on time-average factor prices of countries, the estimated elasticity is larger. This is in line with the view that long-lived factor price changes translate much stronger into changes in technology than short lived ones. This idea forms also the basis of our panel IV estimation. We instrument relative factor prices by taxes. The tax series is constructed from the World Tax Database available at `http://www.bus.umich.edu/otpr/otpr/default.asp`. We use corporate tax rates because they should affect equilibrium interest

[12] Alternative estimation strategies in the literature exploit cointegration properties, cross-country variation in factor price trends, or low-pass filters; see Chirinko (2008).

rates without a strong direct impact on effective relative factor prices when capital can be rented or debt financed. Importantly, our instrumental variable is highly persistent and thus allows us to capture movements in factor prices that are long-lived. As a result, our IV estimator should be closer to the long-run elasticity of substitution. In fact, the estimated 1.28 is almost twice the OLS estimate and thrice the within-country estimate even though we also control for country fixed effects here, too. Together with the fact, that capital-intensity dispersion is higher among plants with more market power, this finding supports the idea of substantive adjustment costs to technology.

Table 5: Estimation of long-run elasticity of substitution

	Dependent variable: $\log(K/L)$			
	OLS	Within-Country	Between-Country	IV
$\log(W/R)$	0.68^{***}	0.43^{***}	0.74^{***}	1.28^{***}
Constant	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	No	Yes
Country fixed effects	No	Yes	No	Yes
Instrument	No	No	No	Yes
R^2	0.76	0.75	0.77	0.71
Countries	99	99	99	99
Observations	2,609	2,609	99	2,609

Notes: Regressions based on unbalanced panel of 99 countries for the years 1956-2002. The instrument is the country-level top marginal income tax rate on domestic corporations. The first stage regression is significant at the 1% level. */**/*** denote 10/5/1% significance.

5 Conclusion

This paper documents a series of new facts about factor misallocation in Chile, Colombia, Germany, and Indonesia. We show that misallocation at the firm/plant level is persistent. We show that it is useful to understand misallocation as being composed of a technological component (the capital intensity, i.e., how things are produced) and a markup component (deviations from optimal size, i.e., how much a firm produces). We show that these two components are roughly orthogonal.

Differences in capital intensities are most important in explaining factor productivity dispersions and they are mostly persistent. At the same time, markup differences across firms are more important in terms of the productivity losses they generate. Markup dispersions show both a strong persistent and a strong transitory component. Finally, we have argued that competition and misallocation are intertwined in a complicated fashion. An increase in competition leads to less technological mislallocation but might create productivity losses from larger markup dispersion if the increase in competition is not uniform across firms.

For future work it would be important to explore whether a dynamic model of technology choice is able to explain our empirical results not only qualitatively but also quantitatively.

References

Asker, J., Collard-Wexler, A., and De Loecker, J. (2014). Dynamic inputs and resource (mis)allocation. *Journal of Political Economy*, 122(5):1013 – 1063.

Asker, J., Collard-Wexler, A., and De Loecker, J. (2017). Market power and production (mis) allocation. a study of the world oil market. Technical report, mimeo.

Bachmann, R. and Bayer, C. (2013). Wait-and-see business cycles? *Journal of Monetary Economics*, 60(6):704 – 719.

Bachmann, R. and Bayer, C. (2014). Investment dispersion and the business cycle. *American Economic Review*, 104(4):1392–1416.

Banerjee, A. V. and Duflo, E. (2005). Growth Theory through the Lens of Development Economics. In Aghion, P. and Durlauf, S., editors, *Handbook of Economic Growth*, volume 1, chapter 7, pages 473–552. Elsevier.

Bar-Isaac, H., Caruana, G., and Cunat, V. (2012). Search, design, and market structure. *American Economic Review*, 102(2):1140–60.

Blalock, G. and Gertler, P. J. (2009). How firm capabilities affect who benefits from foreign technology. *Journal of Development Economics*, 90(2):192–199.

Calvo, G. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398.

Chirinko, B. (2008). [sigma]: The long and short of it. *Journal of Macroeconomics*, 30(2):671–686.

Chirinko, R. S. and Mallick, D. (2017). The Substitution Elasticity, Factor Shares, and the Low-Frequency Panel Model. *American Economic Journal: Macroeconomics*, 9(4):225–253.

David, J., Schmid, L., and Zeke, D. (2018). Risk-adjusted capital allocation and misallocation. *Centre for Economic Policy Research*.

Eslava, M., Haltiwanger, J., Kugler, A., and Kugler, M. (2013). Trade and market selection: Evidence from manufacturing plants in colombia. *Review of Economic Dynamics*, 16(1):135 – 158. Special issue: Misallocation and Productivity.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The Next Generation of the Penn World Table. *American Economic Review*, 105(10):3150–82.

Gilchrist, S. and Williams, J. C. (2000). Putty-Clay and Investment: A Business Cycle Analysis. *Journal of Political Economy*, 108(5):928–960.

Gilchrist, S. and Williams, J. C. (2005). Investment, Capacity, and Uncertainty: A Putty-Clay Approach. *Review of Economic Dynamics*, 8(1):1–27.

Gopinath, G., Kalemli-Ozcan, S., Karabarbounis, L., and Villegas-Sanchez, C. (2017). Capital Allocation and Productivity in South Europe. *Quarterly Journal of Economics (forthcoming)*.

Gourio, F. (2011). Putty-clay technology and stock market volatility. *Journal of Monetary Economics*, 58(2):117–131.

Gourio, F. and Rudanko, L. (2014). Customer Capital. *Review of Economic Studies*, 81(3):1102–1136.

Haltiwanger, J., Kulick, R., and Syverson, C. (2018). Misallocation measures: The distortion that ate the residual. Working Paper 24199, National Bureau of Economic Research.

Henriquez, C. (2008). Stock de Capital en Chile (1985-2005): Metodologia y Resultados. Technical report, Central Bank of Chile.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Johansen, L. (1959). Substitution versus Fixed Production Coefficients in the Theory of Economic Growth: A Synthesis. *Econometrica*, 27(2):pp. 157–176.

Jones, C. (2016). Chapter 1 - the facts of economic growth. volume 2 of *Handbook of Macroeconomics*, pages 3 – 69. Elsevier.

Kaboski, J. P. (2005). Factor price uncertainty, technology choice and investment delay. *Journal of Economic Dynamics and Control*, 29(3):509–527.

Karabarbounis, L. and Neiman, B. (2013). The global decline of the labor share. *Quarterly Journal of Economics*.

Merz, M. and Yashiv, E. (2007). Labor and the Market Value of the Firm. *American Economic Review*, 97(4):1419–1431.

Midrigan, V. and Xu, D. Y. (2014). Finance and misallocation: Evidence from plant-level data. *American Economic Review*, 104(2):422–58.

Oberfield, E. and Raval, D. (2014). Micro Data and Macro Technology. Working paper, Princeton University.

Peters, M. (2016). Heterogeneous Mark-Ups and Endogenous Misallocation. mimeo, Yale University.

Piketty, T. (2011). On the long-run evolution of inheritance: France 1820 2050. *Quarterly Journal of Economics*, 126(3):1071–1131.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

Ramirez Verdugo, A. (2005). Tax Incentives and Business Investment: New Evidence from Mexico. MPRA Paper 2272, University Library of Munich, Germany.

Raval, D. (2014). The Micro Elasticity of Substitution and Non Neutral Technology. Working paper, Federal Trade Commission.

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

Restuccia, D. and Santaeulalia-Llopis, R. (2017). Land misallocation and productivity. Working Paper 23128, National Bureau of Economic Research.

Sheshinski, E. and Weiss, Y. (1977). Inflation and costs of price adjustment. *Review of Economic Studies*, 44(2):287–303.

Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70(1):65–94.

Uras, B. R. and Wang, P. (2017). Production flexibility, misallocation and total factor productivity. Working Paper 23970, National Bureau of Economic Research.

Vial, V. (2006). New Estimates of Total Factor Productivity Growth in Indonesian Manufacturing. *Bulletin of Indonesian Economic Studies*, 42(3):357–369.

Wooford, M. (2011). *Interest and Prices*. Princeton University Press.

A Appendix

A.1 Data description

German Firm Data: USTAN (Unternehmensbilanzstatistiken)

USTAN is itself a byproduct of the Bundesbank's rediscounting and lending activity. The Bundesbank had to assess the creditworthiness of all parties backing promissory notes or bills of exchange put up for rediscounting (i.e., as collateral for overnight lending). It implemented this regulation by requiring the balance-sheet data of all parties involved, which were then archived and collected; see Bachmann and Bayer (2013) for details. Our initial sample consists of 1,846,473 firm-year observations. We remove observations from East German firms to avoid a break in the series in 1990. Finally, we drop the following sectors: hospitality (hotels and restaurants), financial and insurance institutions, public health and education. The resulting sample covers roughly 70% of the West German real gross value added in the private non-financial business sector. In particular, it includes Agriculture, Energy and Mining, Manufacturing, Construction, and Trade.

Chilean Plant Data: ENIA (Encuesta Nacional Industrial Anual)

ENIA is collected by the National Institute of Statistics (*Instituto Nacional de Estadísticas*, INE) and provides plant-level data from 1995 to 2007. ENIA contains information for all manufacturing plants with total employment of at least ten. For the period under analysis, we have a sample of 70,217 plant-year observations. According to INE, this sample covers about 50% of total manufacturing employment.

Colombian Plant Data: EAM (Encuesta Anual Manufacturera)

EAM is a plant-level survey conducted by the National Institute of Statistics (*Departamento Administrativo Nacional de Estaditicas*, DANE) for the period 1977 to 1991.

The survey covers information for all manufacturing plants during 1977-1982; it only contains data on plants with more than ten employees for 1983-1984, and from 1985, small plants are included in small proportion. This results in more than 100,000 plant-year observations.

Indonesian Plant Data: IBS (Survei Tahunan Perusahaan Industri Pengolahan)

IBS is the Indonesian Manufacturing Survey of Large and Medium Establishments, provided by the National Institute of Statistics (*Badan Pusat Statistik*, BPS). The survey covers all plants with 20 or more employees in the manufacturing sector. Given that the capital stock is reported from 1988 onward, we exclude earlier years and focus on the period 1988-2010, with more than 550,000 plant-year observations.

A.2 Sample selection

Starting from the raw data sets, we concentrate on describing the general cleaning steps common to all countries, and we provide more information about country-specific cleaning steps in Table 6.

To begin with, we remove observations where firms or plants report extraordinarily large depreciation rates (e.g., due to fire or accident). The reason is that our dynamic model does not capture such cases, and the perpetual inventory method (PIM) will inaccurately measure the actual capital stock after such incidents occur.[13] Next, for those countries where current values of capital stock are not provided (Germany and Colombia), we recompute capital stocks using the PIM. In conducting the PIM, we drop a small amount of outliers, as explained in Section A.4. Further, we do not consider observations where value-added, capital stock, or employment is non-positive or missing.

Moreover, we do not consider observations where firms/plants have missing values in changes of employment (L), real capital (K) and real value-added (VA).[14] To construct capital productivity, we use the lagged value of capital stock, so we effectively discard the first year of each micro unit. We remove outliers in the levels and in the relative changes of employment, capital, value-added, and factor shares based on 3 standard

[13]In some cases in the ENIA, EAM, and IBS surveys, plants do not report depreciation conditional on positive capital stock. In order to not lose these observations, we impute the depreciation by capital type and two-digit industry, estimating a random effect model, using as the explanatory variable the log-capital stock. To discard rare depreciation events, we drop observations whenever the reported depreciation rate in structures (equipment) is above 40% (60%) yearly. Additionally, we do not consider those cases where the reported depreciation is below 0.1% (1%) in structures (equipment), yearly.

[14]To construct measures of the real capital stock, we consider an index price by each capital type (when available) using the information of gross fixed capital formation at current and constant prices from the National Accounts, while, for value added, we use the GDP price deflator.

deviations from the industry-year mean. In addition, we drop firm-/plant-year observations whenever the total factor expenditures share is either below 1/3 or above 3/2, and whenever the firm/plant average total factor expenditure share is above 1. These two cleaning steps should exclude from our analysis units which continuously report unreasonably large markups or losses.

Finally, as our empirical results rely on a 5-year moving average filter, we do not consider firm-/plant-year observations that have less than 5 consecutive years.

Table 6: Sample selection

Criterion/Country	Germany	Chile	Colombia	Indonesia
Initial sample	1,846,473	70,217	103,006	561,539
East Germany	-115,201	–	–	–
Additional cleaning steps	–	–	–	-32,618
Imputation capital stock	–	–	–	+37,341
Rare depreciation events	-54,280	-8,197	-6,176	-8,775
Outliers in PIM	-73,784	–	-3,960	–
Missing values	-422,739	-21,813	-29,664	-303,663
Outliers in factor variables	-176,232	-5,241	-10,343	-47,940
Less than 5 consecutive years	-312,452	-16,459	-20,338	-93,440
Final sample	689,665	18,507	32,525	112,440

Notes: Missing values summarize the missing values of log value added, log capital, factor shares and log changes in employment, capital and value added. Outliers in factor variables are the sum of all identified outliers at log changes in employment, real capital and real value added, and factor shares. For more information with respect to *Additional cleaning steps* and *Imputation of capital stock* in Indonesia, see Section A.3.

A.3 Indonesia-specific adjustments

Before proceeding with the general cleaning steps applied to all data sets, we need to implement some specific corrections to the Indonesian micro-data. In doing so, we closely follow Blalock and Gertler (2009). First, we correct for mistakes due to data keypunching. If the sum of the capital categories is a multiple of 10^n (with n being an integer) of the total reported capital, we replace the latter with the sum of the categories. Second, we drop duplicate observations within the year (i.e., observations that have the same values for all variables in the survey but differ in their plant identification number). Third, we re-compute value added whenever their values are not consistent with the formula provided by the BPS. Finally, the survey changed its industry classification from ISIC Rev. 2 in 1998 to ISIC Rev. 3 in 1999 and to ISIC Rev. 4 in 2010. We use United

Nations concordance tables to construct a consistent four-digit industry classification.

Further, the surveys from 1996 and 2006 provide information only on the aggregate capital stock, yet not disaggregated by capital type (structure and equipment). To construct an economically reasonable estimate of these variables for these years, we use the average reported investment share and capital share of capital type in the preceding and subsequent year, and impute it, multiplying the aggregate capital stock and investment with the respective share.

Finally, we impute capital stock for plants whenever the survey presents missing values for this variable in plants that reported information in previous and/or subsequent years. Following Vial (2006), we impute capital by type (machinery, vehicles, land and buildings), using the following regression by two-digit sectoral level:

$$\log K_{it} = \beta_0 + \beta_1 \log K_{it-1} + \theta \ln X_{it-1} + \mu_i + \epsilon_{it}$$

where K_{it} is the capital stock of type i, μ_i are plant fixed effects and X_{it-1} is a set of explanatory variables (total output, input, employees, wages, fuel costs and expenditures on materials, leasing, industrial services and taxes).[15]

A.4 Perpetual inventory method

Whenever the data set does not directly provide information on a firm's/plant's capital stock at current values (USTAN and EAM), we re-calculate capital stocks using the perpetual inventory method (PIM), in order to obtain economically meaningful capital series. In doing so, we follow Bachmann and Bayer (2014). To begin with, we compute nominal investment series using the accumulation identity for capital stocks:

$$p_t^I I_{i,k,t} = K^r_{i,k,t+1} - K^r_{i,k,t} + D^r_{i,k,t},$$

where $K^r_{i,k,t}$ and $D^r_{i,k,t}$ are firm/plant i's reported capital stock and depreciation for capital type k at time t, respectively. Given that capital is reported at historical prices and does not reflect the productive (real) level of capital stock, we apply the PIM to construct economic real capital stock at each type of capital:

$$K_{i,k,1} = \frac{p_1^I}{p_{base}^I} K^a_{i,k,1}; \quad K_{i,k,t+1} = K_{i,k,t}(1 - \delta_{i,k,t}) + \frac{p_t^I}{p_{base}^I} I_{i,k,t}, \;\; \forall t \in [0, T]$$

[15] We evaluate the robustness of the imputation procedure using linear interpolation as an alternative approach. Our empirical findings are robust to this alternative specification.

where $K^a_{i,k,1}$ is the accounting value of the capital stock of type k for the first period in which we observe the unit, $\frac{p_t}{p_{base}} I_{i,k,t}$ is the real investment in capital k of firm/plant i at time t and $\delta_{i,k,t}$ is the reported depreciation rate of capital k by firm/plant i at time t.[16]

Even though the aforementioned procedure makes sure that values follow an economically meaningful real capital stock series from the second period onward, it is not clear whether the starting (accounting) input of capital at the unit, $K^a_{i,k,t}$, reflects the productive real value. To account for and adjust the first-period value of capital, we use an iterative approach. Specifically, we construct a time average factor ϕ_k for each type of capital. In the first iteration step, the adjustment factor takes a value of 1, while capital is equal to its balance-sheet value. That is, $K^n_{i,k,t} = \frac{p^I_t}{p^I_{base}} K^a_{i,k,1}$ for $n = 1$. For the subsequent iterations, capital is computed using PIM:

$$K^n_{i,k,t+1} = K^n_{i,k,t}(1 - \delta_{i,k,t}) + \frac{p_t}{p_{base}} I_{i,k,t},$$

while the ajdustment factor is constructed using the ratio between the capital of consecutive iterations

$$\phi^n_k = \frac{1}{NT} \sum_{i,t} \frac{K^n_{i,k,t}}{K^{n-1}_{i,k,t}}.$$

Finally, the capital stock for the first period in which we observe the unit is adjusted by the factor ϕ^n_k. We apply the procedure iteratively until ϕ_k converges[17]

$$K^n_{i,k,1} = \phi^{n-1}_k K^{n-1}_{i,k,1}.$$

[16]The reported depreciation rate is adjusted such that, on average, it coincides with the economic depreciation rate given by the National Accounts. To deflate the investment series, we compute an investment good price deflator from each country using the information of gross fixed capital formation at current and constant prices from the National Accounts.

[17]We stop whenever the value of ϕ_k is below 1.1. At each iteration step, we drop 0.1% from the bottom and top of the capital distribution. This cleaning step makes sure that we do not consider episodes of extraordinary depreciation at the plant, which implies that reported depreciation rates (adjusted to have the same average value from the National Accounts) do not reflect the capital stock given by the PIM.

A.5 Robustness of empirical results

Table 7: Robustness: Transitory and persistent components (HP filtered) of factor productivities, markups, and capital intensity

	$\mathrm{std}(\hat{\alpha}_{it}^L)$	$\mathrm{std}(\hat{\alpha}_{it}^K)$	$\rho(\hat{\alpha}_{it}^L, \hat{\alpha}_{it}^K)$	$\mathrm{std}(\bar{\alpha}_{it}^L)$	$\mathrm{std}(\bar{\alpha}_{it}^K)$	$\rho(\bar{\alpha}_{it}^L, \bar{\alpha}_{it}^K)$
	Transitory Component (HP)			Persistent Component (HP)		
DE	0.062	0.113	0.352	0.236	0.471	-0.223
CL	0.208	0.286	0.542	0.293	0.614	-0.078
CO	0.142	0.168	0.551	0.273	0.682	-0.074
ID	0.267	0.386	0.483	0.354	0.737	0.013
	$\mathrm{std}(\hat{\mu}_{it})$	$\mathrm{std}(\hat{\kappa}_{it})$	$\rho(\hat{\mu}_{it}, \hat{\kappa}_{it})$	$\mathrm{std}(\bar{\mu}_{it})$	$\mathrm{std}(\bar{\kappa}_{it})$	$\rho(\bar{\mu}_{it}, \bar{\kappa}_{it})$
	Transitory Component (HP)			Persistent Component (HP)		
DE	0.060	0.108	-0.157	0.175	0.572	0.055
CL	0.202	0.246	-0.078	0.240	0.701	-0.088
CO	0.135	0.149	-0.080	0.288	0.753	-0.507
ID	0.258	0.347	-0.100	0.310	0.813	-0.101

Notes: Labor productivity, a_{it}^L, and capital productivity, a_{it}^K, as defined in (1). Markups, μ_{it}, and capital intensity, κ_{it}, as defined in (3) and (4). HP: results based on the decomposing between transitory and persistent using an HP filter ($\lambda = 6.25$). Factor productivities are demeaned by 4-digit industry and year and expressed in logs. Standard errors are clustered standard errors at the firm/plant level. ρ denotes correlation. DE: Germany, CL: Chile, CO: Colombia, ID: Indonesia.

Table 8: Robustness: Weighted second moments of factor productivities, markups, and capital intensity at different frequencies

	std($\hat{\alpha}^L_{it}$)	std($\hat{\alpha}^K_{it}$)	$\rho(\hat{\alpha}^L_{it}, \hat{\alpha}^K_{it})$	std($\bar{\alpha}^L_{it}$)	std($\bar{\alpha}^K_{it}$)	$\rho(\bar{\alpha}^L_{it}, \bar{\alpha}^K_{it})$
	Transitory Component (5Y MA)			Persistent Component (5Y MA)		
DE	0.050	0.101	0.316	0.196	0.457	-0.176
CL	0.230	0.311	0.553	0.312	0.584	-0.085
CO	0.152	0.180	0.555	0.278	0.674	-0.076
ID	0.298	0.420	0.494	0.368	0.741	0.015
	std($\hat{\mu}_{it}$)	std($\hat{\kappa}_{it}$)	$\rho(\hat{\mu}_{it}, \hat{\kappa}_{it})$	std($\bar{\mu}_{it}$)	std($\bar{\kappa}_{it}$)	$\rho(\bar{\mu}_{it}, \bar{\kappa}_{it})$
	Transitory Component (5Y MA)			Persistent Component (5Y MA)		
DE	0.052	0.090	-0.161	0.172	0.503	0.067
CL	0.221	0.269	-0.082	0.241	0.690	-0.086
CO	0.144	0.159	-0.086	0.289	0.748	-0.513
ID	0.288	0.373	-0.101	0.321	0.821	-0.104

Notes: Labor productivity, a^L_{it}, and capital productivity, a^K_{it}, as defined in (1). Markups, μ_{it}, and capital intensity, κ_{it}, as defined in (3) and (4). Cross-sectional standard deviations (std) and correlation (ρ) of transitory and persistent components. Transitory and persistent components are obtained by applying a five year moving average filter (5Y MA). Moments are weighted based on the value-added of the plant/firm. Variables under interest are demeaned by 4-digit industry and year and expressed in logs. Standard errors in parentheses are clustered standard errors at the firm/plant level. DE: Germany, CL: Chile, CO: Colombia, ID: Indonesia.

Table 9: Robustness: Transitory and persistent components (9 year moving average filter) of factor productivities, markups, and capital intensity

	std($\hat{\alpha}_{it}^L$)	std($\hat{\alpha}_{it}^K$)	$\rho(\hat{\alpha}_{it}^L, \hat{\alpha}_{it}^K)$	std($\bar{\alpha}_{it}^L$)	std($\bar{\alpha}_{it}^K$)	$\rho(\bar{\alpha}_{it}^L, \bar{\alpha}_{it}^K)$
	Transitory Component (9YMA)			Persistent Component (9YMA)		
DE	0.074	0.140	0.350	0.204	0.406	-0.203
CL	0.230	0.345	0.479	0.229	0.521	-0.111
CO	0.165	0.211	0.499	0.237	0.604	-0.079
ID	0.297	0.466	0.445	0.287	0.644	-0.046
	std($\hat{\mu}_{it}$)	std($\hat{\kappa}_{it}$)	$\rho(\hat{\mu}_{it}, \hat{\kappa}_{it})$	std($\bar{\mu}_{it}$)	std($\bar{\kappa}_{it}$)	$\rho(\bar{\mu}_{it}, \bar{\kappa}_{it})$
	Transitory Component (9YMA)			Persistent Component (9YMA)		
DE	0.073	0.134	-0.184	0.157	0.490	0.089
CL	0.225	0.309	-0.111	0.193	0.593	-0.106
CO	0.158	0.193	-0.122	0.248	0.667	-0.487
ID	0.292	0.427	-0.130	0.253	0.716	-0.100

Notes: Labor productivity, a_{it}^L, and capital productivity, a_{it}^K, as defined in (1). Markups, μ_{it}, and capital intensity, κ_{it}, as defined in (3) and (4). 9YMA: results based on the decomposing between transitory and persistent using a 9 year moving average filter. Factor productivities are demeaned by 4-digit industry and year and expressed in logs. Standard errors are clustered standard errors at the firm/plant level. ρ denotes correlation. DE: Germany, CL: Chile, CO: Colombia, ID: Indonesia.

Table 10: Robustness: Dispersion of capital intensity and markups

	DE	CL	CO	ID
	$\epsilon^2_{\bar{\kappa}_{it}}$			
Log-Markup	0.024	0.042	0.097	0.033
	(0.003)	(0.016)	(0.014)	(0.010)
Log-Size	-0.026	-0.072	-0.036	0.011
	(0.003)	(0.018)	(0.018)	(0.012)

Notes: The results are obtained based on a two-step procedure. First, we remove cross-sectional differences in log capital intensity (κ_{it}) that can be explained by the log of markups and size. Second, the squared estimated residual based on the first stage, ($\epsilon^2_{\bar{\kappa}_{it}}$), is regressed on the standarized log of markups and size. Standard errors in parentheses are clustered standard errors at the firm/plant level. DE: Germany, CL: Chile, CO: Colombia, ID: Indonesia.

A.6 Data treatment and dispersions

Table 11: Effects of data treatment steps on log-dispersions: Chile

	TFPR	Markups	Technology
	$\frac{(py)_{it}}{(R_{it}K_{it})^{\alpha}(W_tL_{it})^{1-\alpha}}$	$\frac{(py)_{it}}{R_{it}K_{it}+W_tL_{it}}$	$\frac{R_{it}K_{it}}{W_tL_{it}}$
	[Ignore differences in capital composition]		
Initial sample	-	-	-
	[0.662]	[0.595]	[1.227]
(1) Outliers in δ	-	-	-
	[0.622]	[0.575]	[1.086]
(2) Outliers in markups	0.450	0.421	0.855
	[0.466]	[0.409]	[1.035]
(3) Outliers in growth of K, L, PY	0.430	0.405	0.820
	[0.443]	[0.391]	[0.989]
(4) Outliers in α^L and α^K	0.393	0.362	0.802
	[0.405]	[0.346]	[0.967]
(5) Five consecutive years	0.348	0.322	0.713
	[0.378]	[0.321]	[0.906]

Notes: All moments are computed after trimming all log TFPR observations above the 99th percentile and below the 1st percentile. We set $\alpha = 1/3$. In brackets, we present log dispersions when ignoring firm-/plant-year variation in composition of the capital stock across structures and equipment. In practice, that means we set the user cost of capital uniformly to 15% and conduct the PIM for aggregate unit-specific capital stock series.

Table 12: Effects of data treatment steps on log-dispersions: Colombia

	TFPR	Markups	Technology
	$\frac{(py)_{it}}{(R_{it}K_{it})^{\alpha}(W_t L_{it})^{1-\alpha}}$	$\frac{(py)_{it}}{R_{it}K_{it}+W_t L_{it}}$	$\frac{R_{it}K_{it}}{W_t L_{it}}$
	[Ignore differences in capital composition]		
Initial sample	-	-	-
	[0.494]	[0.439]	[1.082]
(1) Outliers in δ	-	-	-
	[0.486]	[0.437]	[1.065]
(2) Adjust K by PIM	0.456	0.511	0.899
	[0.471]	[0.516]	[0.960]
(3) Outliers in markups	0.363	0.395	0.861
	[0.368]	[0.387]	[0.892]
(4) Outliers in growth of K, L, PY	0.355	0.386	0.844
	[0.358]	[0.378]	[0.864]
(5) Outliers in α^L and α^K	0.339	0.357	0.817
	[0.341]	[0.346]	[0.833]
(6) Five consecutive years	0.312	0.324	0.739
	[0.323]	[0.325]	[0.772]

See notes of Table 11 for details.

Table 13: Effects of data treatment steps on log-dispersions: Indonesia

	TFPR	Markups	Technology
	$\frac{(py)_{it}}{(R_{it}K_{it})^{\alpha}(W_t L_{it})^{1-\alpha}}$	$\frac{(py)_{it}}{R_{it}K_{it}+W_t L_{it}}$	$\frac{R_{it}K_{it}}{W_t L_{it}}$
	[Ignore differences in capital composition]		
Initial sample	-	-	-
	[0.718]	[0.699]	[1.134]
(1) Outliers in δ	-	-	-
	[0.707]	[0.686]	[1.128]
(2) Outliers in markups	0.529	0.495	1.031
	[0.528]	[0.496]	[1.048]
(3) Outliers in growth of K, L, PY	0.516	0.487	0.987
	[0.509]	[0.484]	[0.983]
(4) Outliers in α^L and α^K	0.500	0.463	0.955
	[0.493]	[0.459]	[0.950]
(5) Five consecutive years	0.453	0.424	0.869
	[0.461]	[0.432]	[0.890]

See notes of Table 11 for details.

A.7 Second-order approximation of aggregate productivity

Note that aggregate costs are given by $C = \int c_i y_i di$. Average costs, C/Y, defined as aggregate costs relative to aggregate output, is an (inverse) measure of aggregate TFP. Taking aggregate output Y as given and using (Assumption 3) and equation (6), aggregate costs are

$$C = Y \left[\int c(k)^{1-\eta} z^{\eta-1} m^{-\eta} dF(k, z, m) \right] \left[\int c(k)^{1-\eta} z^{\eta-1} m^{1-\eta} dF(k, z, m) \right]^{\frac{\eta}{1-\eta}}, \tag{8}$$

where F is the joint distribution of firms across technology, demand shock, and markup. The last term captures the aggregate price index. Aggregate unit costs, C/Y, are equivalent to aggregate productivity, and after some reformulation, we obtain

$$\begin{aligned}\log\left(\frac{C}{Y}\right) &= -\log(\bar{z}) + \underbrace{\log\left[\int c(k)^{1-\eta} \tilde{z}^{\eta-1} \tilde{m}^{-\eta} dF(k, \tilde{z}, \tilde{m})\right]}_{g^c} \\ &\quad + \frac{\eta}{1-\eta} \underbrace{\log\left[\int c(k)^{1-\eta} \tilde{z}^{\eta-1} \tilde{m}^{1-\eta} dF(k, \tilde{z}, \tilde{m})\right]}_{g^p},\end{aligned} \tag{9}$$

where $\log(\bar{z}) = \mathbb{E}\log(z)$, $\tilde{z} = z/\bar{z}$, $\log(\bar{m}) = \mathbb{E}\log(m)$, and $\tilde{m} = m/\bar{m}$. To characterize the effects of dispersion in technology, the demand shock, and the markup, we provide a second-order approximation in log deviations. We define $\kappa = \log k$, $\tilde{c}(\kappa) = \log(W + R\exp\kappa) - \log(f(\exp\kappa))$, $\tilde{\zeta} = \log\tilde{z}$, and $\tilde{\mu} = \log\tilde{m}$. The approximation will be around the cost-minimizing capital intensity $\kappa^* = \arg\min_\kappa \tilde{c}(\kappa)$, a unit demand shock $\zeta = \log z = 0$, and a unit markup $\mu = \log m = 0$. Define

$$\begin{aligned} g^c &= \log \int \exp\{h^c\} dF, \\ h^c &= (1-\eta)\tilde{c}(\kappa^* + \sigma(\kappa - \kappa^*)) + (\eta - 1)\sigma\tilde{\zeta} - \eta\sigma\tilde{\mu}, \end{aligned}$$

where we perturb $(\kappa, \tilde{\zeta}, \tilde{\mu})$ by σ. Evaluated at $\sigma = 0$, the first and second derivative of h^c w.r.t. σ are

$$\begin{aligned} \frac{\partial h^c}{\partial\sigma}\Big|_{\sigma=0} &= (1-\eta)\underbrace{\tilde{c}'(\kappa^*)}_{=0}(\kappa - \kappa^*) + (\eta-1)\tilde{\zeta} - \eta\tilde{\mu}, \\ \frac{\partial^2 h^c}{(\partial\sigma)^2}\Big|_{\sigma=0} &= (1-\eta)\tilde{c}''(\kappa^*)(\kappa - \kappa^*)^2. \end{aligned}$$

Note that $\sigma = 0$ implies $\exp\{-g^c\}\exp\{h^c\} = 1$, which we use to obtain

$$\frac{\partial g^c}{\partial \sigma}\Big|_{\sigma=0} = \int [(\eta-1)\tilde{\zeta} - \eta\tilde{\mu}]dF = 0,$$
$$\frac{\partial^2 g^c}{(\partial \sigma)^2}\Big|_{\sigma=0} = (\eta-1)^2\mathbb{V}(\zeta) - \eta(\eta-1)\mathrm{Cov}(\zeta,\mu) + \eta^2\mathbb{V}(\mu) - (\eta-1)\tilde{c}''(\kappa^*)\mathbb{E}[(\kappa-\kappa^*)^2].$$

Analogously, we obtain

$$\frac{\partial g^p}{\partial \sigma}\Big|_{\sigma=0} = \int [(\eta-1)\tilde{\zeta} + (1-\eta)\tilde{\mu}]dF = 0,$$
$$\frac{\partial^2 g^p}{(\partial \sigma)^2}\Big|_{\sigma=0} = (\eta-1)^2\mathbb{V}(\zeta) - (\eta-1)^2\mathrm{Cov}(\zeta,\mu) + (\eta-1)^2\mathbb{V}(\mu) - (\eta-1)\tilde{c}''(\kappa^*)\mathbb{E}[(\kappa-\kappa^*)^2].$$

A second-order Taylor approximation of average unit costs yields

$$\begin{aligned}\log\left(\frac{C}{Y}\right) - \log\left(\frac{C}{Y}\right)^* &\approx -\mathbb{E}[\zeta] + \frac{\partial^2 g^c}{(\partial \sigma)^2}\Big|_{\sigma=0} - \frac{\eta}{\eta-1}\frac{\partial^2 g^p}{(\partial \sigma)^2}\Big|_{\sigma=0} \\ &= -\mathbb{E}[\zeta] - (\eta-1)\mathbb{V}(\zeta) + \eta\mathbb{V}(\mu) + \tilde{c}''(\kappa^*)\mathbb{E}[(\kappa-\kappa^*)^2]. \qquad (10)\end{aligned}$$

We further use

$$\tilde{c}''(\kappa^*) = \frac{\partial^2 \log c(k)}{(\partial \log k)^2}\Big|_{k=k^*} = \frac{1}{2\rho}s^*(1-s^*),$$

where $s^* = \frac{Rk^*}{W+Rk^*}$ is the capital expenditure share in the cost-minimizing optimum. In addition, we approximate $\mathbb{E}[(\kappa-\kappa^*)^2]$ by $\mathbb{V}(\kappa)$, which will bias downward the importance of technology dispersions.

A.8 Aggregate productivity and market structure

We propose a stylized model that offers important insights: it shows that while an increase in markup dispersion leads to a ceteris paribus decrease in productivity (and welfare), an increased markup dispersion might be an epiphenomenon of an otherwise welfare-improving change. Note, however, that this does not mean that the economy could not be better off without markup dispersion. The first-best net markup is obviously always zero on all goods, and thus, the first-best markup dispersion is always zero.

We consider the following market structure: Each firm produces a variety i as in our baseline setup. We assume that the unit costs of production are normalized to unity. The firm might face a competitor who produces a perfect substitute, with probability λ_C and who also has costs of production equal to one. With probability λ_I the firm innovates and increases its productivity, such that the unit costs of production are now

$\underline{c} < 1$. We assume that the innovation is drastic, such that $\underline{c}\frac{\eta}{\eta-1} < 1$.

This implies that the firm will set the markup equal to $\bar{\mu} = \frac{\eta}{\eta-1}$ whenever it is not facing a competitor or faces a competitor and has innovated; otherwise, the gross markup is unity. The expected markup and the expected productvity in the economy are therefore

$$\begin{aligned} \mathbb{E}(\mu) &= \lambda_C(1-\lambda_I) + (1-\lambda_C(1-\lambda_I))\bar{\mu} & (11) \\ \mathbb{E}(c) &= (1-\lambda_I) + \lambda_I \underline{c}, & (12) \end{aligned}$$

such that the average markup is increasing in innovation and decreasing in competition, whereas the expected (unweighted) costs are decreasing in innovation. The variance of markups is given by

$$\mathbb{V}(\mu) = \lambda_C(1-\lambda_I)(1-\lambda_C(1-\lambda_I))(\bar{\mu}-1)^2 \qquad (13)$$

and hence increasing in *competition* whenever $\lambda_C(1-\lambda_I) < 1/2$ and increasing in *innovation* whenever $\lambda_C(1-\lambda_I) > 1/2$. In words, when the economy is not very innovative and not very competitive, an increase in competition increases markup dispersions and thus lowers aggregate productivity, but it also lowers average markups. When the economy is very competitive but not very innovative, an increase in innovation increases markup dispersions (and average markups), but it lowers production costs. Note that under $\lambda_C = 0$, firms are monopolistically competitive. In this sense, especially for economies (or sectors therein) that are almost perfectly competitive, i.e. $\lambda_C(1-\lambda_I) > 1/2$, caution is warranted: higher markup dispersion may be a result of fast productivity growth and fast productivity growth may be a result of low average productivity. In that sense, large misallocation may be the consequence and not only the cause of low productivity if developing countries are catching up.