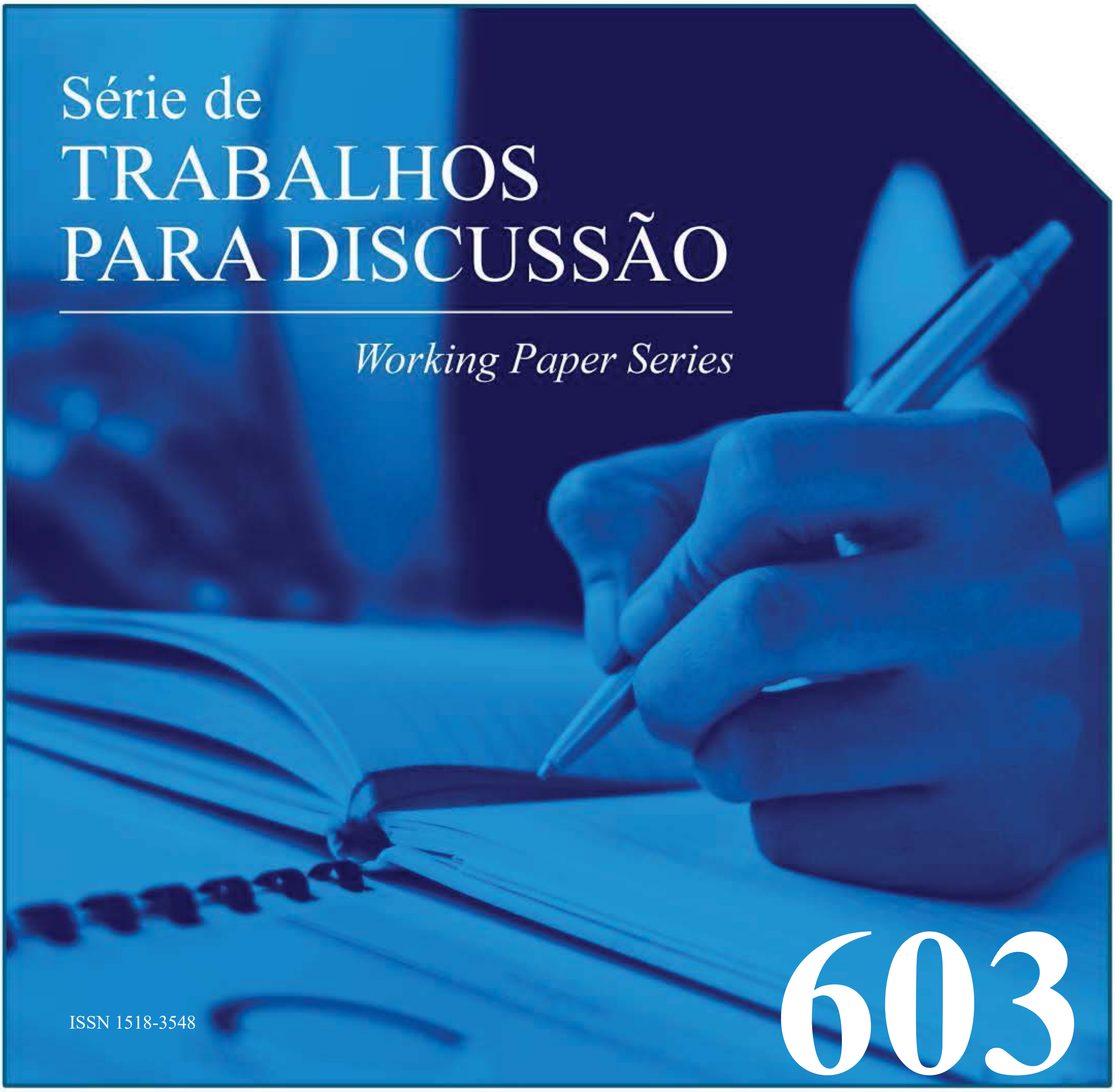

Outubro 2024

Revisiting the Facts of Economic Growth: insights from assessing misallocation over 70 years for up to 100 countries

Tomás R. Martinez, Thiago Trafane Oliveira Santos

BANCO CENTRAL
DO BRASIL

ISSN 1518-3548
CGC 00.038.166/0001-05

Working Paper Series	Brasília	no. 603	Outubro	2024	p. 3-86

Working Paper Series

Edited by the Research Department (Depep) – E-mail: workingpaper@bcb.gov.br

Editor: Rodrigo Barbone Gonzalez

Co-editor: Eurilton Alves Araujo Jr

Head of the Research Department: André Minella

Deputy Governor for Economic Policy: Diogo Abry Guillen

Citizen Service Division

Banco Central do Brasil

Deati/Diate

SBS – Quadra 3 – Bloco B – Edifício-Sede – 2º subsolo

70074-900 Brasília – DF – Brazil

Toll Free: 0800 9792345

Fax: +55 (61) 3414-2553

Internet: http//www.bcb.gov.br/?CONTACTUS

Non-technical Summary

There are two reasons for a country being richer, producing more goods and services: (i) it employs more inputs (land, equipment, workers, etc.) in production, or (ii) it is more efficient in using those inputs. This efficiency level is known as total factor productivity (TFP) and is a key determinant of countries' economic performance. However, TFP is a residual, the portion of production not explained by the inputs. In this context, stating TFP is a crucial determinant of economic development is equivalent to saying we do not understand a significant portion of it. This fact has spurred extensive research into the determinants of TFP. In particular, a literature has emerged arguing that a country with lower TFP does not necessarily have technologically lagging firms. It may be the case that resources are not allocated efficiently across firms due to factors such as taxes, regulations, subsidized credit, corruption, and market frictions. We focus on misallocation due to market power: if more productive firms have increased market power, charging higher prices, they will sell less and consequently be smaller than they ideally should be, thereby lowering the average productivity in the economy, that is, the TFP. The issue is that measuring misallocation typically requires firm-level data available only for selected countries in specific years – and usually restricted to manufacturing. This fact has hindered assessments of the role played by misallocation in shaping TFP. This paper addresses this issue by developing a model that estimates market-power-driven misallocation using primarily macroeconomic data. We apply this framework to decompose TFP into technology and allocative efficiency components from 1950 to 2019 for up to a hundred countries. Utilizing this decomposition, we revisit key facts of economic growth. On the one hand, we evaluate the world income frontier as proxied by the US, finding that changes in misallocation can significantly impact short-run growth. On the other hand, we examine the economic performance around the world. We conclude misallocation enhances our understanding of cross-country income differences, even though a substantial unexplained portion persists. We also find countries are not converging to a common level of allocative efficiency, suggesting misallocation is linked, in the long run, to long-lasting country-specific factors such as institutions.

Sumário Não Técnico

Existem duas razões para um país ser mais rico, produzindo mais bens e serviços: (i) ele emprega mais insumos (terra, máquinas, trabalhadores, etc.) na produção, ou (ii) ele é mais eficiente no uso desses insumos. Esse nível de eficiência é conhecido como produtividade total dos fatores (PTF) e é um determinante fundamental do desempenho econômico dos países. No entanto, a PTF é um resíduo, a parcela da produção não explicada pelos insumos. Nesse contexto, afirmar que a PTF é um determinante crucial do desenvolvimento econômico é equivalente a dizer que não entendemos uma parte significativa desse fenômeno. Esse fato estimulou uma extensa pesquisa sobre os determinantes da PTF. Em particular, uma literatura surgiu argumentando que um país com PTF mais baixa não necessariamente possui empresas tecnologicamente atrasadas. É possível que os recursos não estejam alocados eficientemente entre as firmas devido a fatores como impostos, regulamentações, crédito subsidiado, corrupção e fricções de mercado. Nós focamos na má alocação causada por poder de mercado: se as firmas mais produtivas têm maior poder de mercado, cobrando preços mais altos, elas venderão menos e, logo, serão menores do que deveriam idealmente ser, reduzindo assim a produtividade média, isto é, a PTF. A questão é que medir a má alocação normalmente requer dados ao nível das firmas disponíveis apenas para países selecionados em anos específicos – e geralmente restritos à manufatura. Esse fato limitou as avaliações do impacto da má alocação na PTF. Este artigo supera esse obstáculo ao desenvolver um modelo que estima a má alocação causada por poder de mercado usando principalmente dados macroeconômicos. Aplicamos essa metodologia para decompor a PTF em componentes de tecnologia e eficiência alocativa de 1950 a 2019 para até cem países. Utilizando essa decomposição, reexaminamos importantes fatos do crescimento econômico. Por um lado, avaliamos o desempenho dos países mais ricos, medindo-o nos Estados Unidos. Constatamos que mudanças na eficiência alocativa podem impactar significativamente o crescimento de curto prazo. Por outro lado, examinamos o desempenho econômico ao redor do mundo. Concluímos que a má alocação desempenha um papel significativo na explicação das diferenças de renda entre países, embora uma parte substancial permaneça inexplicada. Também encontramos que os países não estão convergindo para um nível comum de eficiência alocativa, sugerindo que a má alocação está ligada, no longo prazo, a fatores duradouros específicos de cada país, como as instituições.

Revisiting the Facts of Economic Growth: insights from assessing misallocation over 70 years for up to 100 countries*

Tomás R. Martinez
Insper
tomas.martinez@insper.edu.br

Thiago Trafane Oliveira Santos
Banco Central do Brasil
thiago.trafane@bcb.gov.br

Abstract

Assessments of the role played by misallocation in shaping total factor productivity (TFP) have been hindered by constraints in the availability of firm-level data. This paper addresses this issue by developing a static Cournot model that primarily requires standard macroeconomic data to estimate market-power-driven misallocation. We apply this framework to decompose aggregate TFP into technology and allocative efficiency components from 1950 to 2019 for up to a hundred countries from the Penn World Table 10.01. Utilizing this decomposition, we revisit key facts of economic growth. On the one hand, we evaluate the world income frontier as proxied by the US, finding that changes in misallocation can significantly impact short-run growth. On the other hand, we examine the economic performance around the world. We conclude misallocation enhances our understanding of cross-country income differences, even though a substantial unexplained portion persists. We also find a lack of convergence in allocative efficiency, suggesting market-power-driven misallocation is linked, in the long run, to long-lasting country-specific factors such as institutions.

JEL codes: L13, O11, O33, O40, O47.
Keywords: TFP, misallocation, Cournot model, development accounting, growth accounting.

*We are grateful to Stephen L. Parente for his comments and suggestions on earlier versions of this paper.

1 Introduction

Aggregate total factor productivity (TFP) is a key determinant of economic performance across countries and time (e.g., Klenow and Rodríguez-Clare 1997; Caselli 2005, 2016; Jones 2016; Bergeaud et al. 2018; and Crafts and Woltjer 2021). However, TFP is usually regarded as a residual, "a measure of our ignorance," capturing the unexplained portion of economic outcomes. This has led to extensive research into the determinants of TFP, as highlighted by the title of Prescott's (1998) paper: "Needed: A Theory of TFP." Jones (2016, p.46) argues that "the literature on misallocation has emerged to provide the kind of theory that Prescott was seeking." Yet measuring misallocation typically requires firm-level data, available only for selected countries in specific years—and usually limited to the manufacturing sector. This constraint has presented obstacles to including misallocation estimates in standard growth and development accounting exercises. This paper addresses this task by decomposing aggregate TFP into technology and allocative efficiency components from 1950 to 2019, covering up to a hundred countries.

We employ a static model where firms engage in Cournot competition. In the model, (i) a firm's market share increases with its productivity, and (ii) a firm with a larger market share faces lower price elasticity of demand, charging a higher markup. Consequently, variations in productivity across firms result in markup variations, leading to inefficient allocation of resources as the marginal products are not equalized across firms. Our model, therefore, focuses specifically on market-power-driven misallocation, aligning with recent papers that estimate the welfare costs of markups such as Edmond et al. (2022). It is worth noting that this form of allocative inefficiency is typical in Cournot models (e.g., Atkeson and Burstein 2008). The novel aspect of our model lies in its primary reliance on standard macroeconomic data for calibration, which is precisely what enables our comprehensive assessment of misallocation across both time and space.

The minimal microdata requirement does not come without a cost. Our model relies on stronger assumptions than other related oligopoly models, such as Edmond et al. (2015) and De Loecker et al. (2021). First, we assume a single sector where firms produce a homogeneous good, rather than firms producing differentiated goods in each sector over a continuum of sectors. Second, there are no fixed costs. However, since the goods are homogeneous, some firms may still be inactive, and the set of active firms is determined through an entry stage. Third, we assume free entry among low-productivity firms, with the number of inefficient firms being sufficiently large that some remain inactive. As a result, the profit of the marginal active firm should be insignificant.

We show allocative efficiency and other key model expressions are independent of parameters such as the price elasticity of demand and the number of firms, requiring only the empirical distribution of active firms' productivity. This feature is crucial for decomposing TFP using macroeconomic data and stems from the model's stronger assumptions, notably free entry without fixed costs—a scenario possible only if goods are homogeneous. In such circumstances, the market share of the marginal active firm is negligible, regardless of parameter values. By contrast, when goods are heterogeneous, firm selection requires fixed costs, leading the marginal firm to produce positive quantities in equilibrium, even with free entry. In these cases, the marginal firm's output depends on model parameters (e.g., demand elasticity and fixed costs), as identified from the zero-profit condition.

Therefore, to decompose the TFP, we only need to estimate the empirical distribution of active firms' productivity. Consistent with a large part of the literature (e.g., Melitz and Redding 2015, Edmond et al. 2015, and Edmond et al. 2022), we assume this distribution is Pareto, truncated within the productivity range. We test different values for the shape parameter and estimate the distributional support by targeting (i) the aggregate TFP and (ii) the cost-weighted average of firm-level markups. To compute these moments using real-world data, we parameterized the firms' constant-returns-to-scale Cobb-Douglas production function, which has both physical and human capital as inputs. We employ the standard physical capital share parameter $\alpha = 1/3$ in our baseline calibration. Once α is determined, the two target moments are easily obtained from standard macroeconomic data. The TFP is backed out as a residual in the aggregate production function. The cost-weighted average markup equals $1 - \alpha$ divided by the labor share of national income, aligning with Hall (1988), which has recently gained widespread adoption for assessing firm-level markups (e.g., De Loecker and Warzynski 2012, De Loecker and Eeckhout 2018, De Loecker et al. 2020, Traina 2018, Calligaris et al. 2018, and Autor et al. 2020).

Our oligopoly model accommodates variability in the labor share of national income, interpreting it as informative of misallocation, since the estimation of allocative efficiency relies exclusively on the average markup.[1] Specifically, a higher labor share—and thus a lower average markup—is associated with reduced misallocation, with allocation becoming optimal as the markup converges to one. Intuitively, an average markup closer to the competitive unitary level suggests a less distorted economy and, therefore, greater proximity to optimal allocation. After estimating allocative efficiency from the average markup, we pin down the other element of productivity—the technology component—using TFP data. Hence, in our model, the residual of the production function is not TFP itself but rather the technology component alone, which is cleaner as it is free of misallocation effects.

Employing this calibration strategy, we decompose the aggregate TFP for various countries and years using data from the Penn World Table 10.01 (Feenstra et al. 2015). We use this decomposed TFP data to revisit key facts of economic growth. On the one hand, we evaluate nations at the income frontier, using the United States as a proxy, as done in Jones (2016). We begin by demonstrating that our market-power-driven misallocation estimates are consistent with those obtained from the oligopoly model of Edmond et al. (2022) for the US. This serves

[1]In contrast, under perfect competition, the labor income share would be identical across firms and over time, with inputs always allocated optimally.

as an important validation of our model, particularly considering their model is more flexible, general, requiring firm-level data for calibration. Our empirical analysis, spanning from 1954 to 2019, reveals that changes in misallocation can significantly impact short-run growth. For example, during 2000–2007, the US witnessed notable technological improvement coupled with declining allocative efficiency. This suggests the dot-com boom and information technology (IT) advancements led to productivity gains but concentrated in certain firms. On a more general note, the technology component seems to grow more steadily than the TFP itself, at around 1% per year. Notable exceptions are the periods of 1954–1973 and 2000–2007, when technology contributed approximately 2% annually.

On the other hand, we examine the economic performance around the world. Although the Penn World Table 10.01 provides data from 1950 to 2019, the sample coverage varies across the years, encompassing more than a hundred countries in recent years. Our findings underscore the significant role of misallocation in explaining cross-country income differences. Despite its significance, a considerable unexplained portion persists, still constituting the majority of observed variability in most cases. Additionally, we obtain limited support for the convergence hypothesis in income and both components of TFP. Consequently, countries do not appear to be converging over time to a common degree of allocative efficiency, indicating the level of efficiency is country-specific even in the long run. Interestingly, this suggests market-power-driven misallocation is linked, in the long run, to long-lasting country-specific factors such as institutions.

Related literature. From a methodological perspective, our work is related to several papers that embed oligopoly market structures in macroeconomic models. In most models, firms' decisions are static as in our own (Bernard et al. 2003, Atkeson and Burstein 2008, Edmond et al. 2015, De Loecker et al. 2021), but there are also models in which they are dynamic (Peters 2020, Wang and Werning 2022, Edmond et al. 2022).[2] Besides their methodological similarities, these papers have very different objectives. For instance, Atkeson and Burstein (2008) seek to explain the observed deviations from relative purchasing power parity, Edmond et al. (2015) evaluate the impact of opening up to trade on productivity through misallocation, and Wang and Werning (2022) study how market concentration affects the potency of monetary policy. Among them, Edmond et al. (2022) is the closest to our paper in terms of purposes, as they assess the welfare costs of markups using (also) a Cournot model with free entry. However, their model is dynamic and involves a different form of free entry. While in our model firms decide to enter the market only *after* knowing their productivity, in their model firms should decide *before* knowing it. Furthermore, Edmond et al. (2022) estimate such costs just for the US, not on a period-by-period basis, and consider markup costs associated with factors beyond misallocation, such as inefficient entry. Furthermore, they need firm-level data to calibrate their model. We contribute to this literature by developing a model that requires minimal microdata

[2]Berger et al. (2022) also include strategic behavior in a general equilibrium framework, but they do that by considering an oligopsony model for the labor market.

for calibration, allowing us to assess market-power-driven misallocation for a broader set of countries and years.

Given our goal of revisiting key facts of economic growth, this study closely aligns with the development economics literature, particularly that focused on macroeconomic analyses and development accounting exercises (e.g., Klenow and Rodríguez-Clare 1997; Prescott 1998; Caselli 2005, 2016; and Jones 2016). Jones (2016) serves as our primary reference, offering a comprehensive overview of economic growth facts, some of which we reexamine in this study. Our contribution to this literature lies in providing a comprehensive assessment of misallocation across both time and space. As a consequence, we address allocative efficiency issues within development accounting frameworks for several different years. Furthermore, we provide evidence on the impacts of misallocation on frontier growth, a topic that has received considerably less attention than the study of misallocation and development (Jones 2016).

Our work also shares common ground with the growth-accounting literature that decomposes TFP growth into technology and allocative efficiency components (Basu and Fernald 2002, Petrin and Levinsohn 2012, Baqaee and Farhi 2020).[3] It is most closely related to Baqaee and Farhi (2020) since they, like us, measure allocative efficiency as the distance from the optimal allocation.[4] However, their model is much more general as they do not impose any specific market structure and allow for arbitrary elasticities of substitution, returns to scale, factor mobility, and input-output network linkages. As a result, the quantification of their model requires extensive microdata that are hardly available, while our model requires mainly macroeconomic data.

Finally, our paper is also related to the literature on misallocation (see, e.g., Restuccia and Rogerson 2008, Hsieh and Klenow 2009, and the survey article by Restuccia and Rogerson 2013). However, in this literature, misallocation is a result of exogenous wedges, whose estimation requires extensive firm-level data, similar to growth-accounting methods, especially Baqaee and Farhi (2020).[5] In contrast, misallocation is endogenous in our model, emerging as an equilibrium outcome. Furthermore, in this literature, the typical goal is to gauge the importance of misallocation in explaining cross-country TFP differences, usually within the manufacturing sector due to data availability, while we focus on economy-wide outcomes, both across countries and over time.

[3] Baqaee and Farhi (2020) present a review of such methods, showing they also use different definitions for the relevant aggregate productivity. For a broader review of the growth-accounting literature, see Hulten (2010).

[4] Measuring allocative efficiency as the distance from optimal allocation aligns with the prevailing notion in the misallocation literature. However, alternative concepts exist, particularly in the growth-accounting literature. As pointed out by Baqaee and Farhi (2020, p.107), "the growth-accounting notion of changes in allocative efficiency due to the reallocation of resources to more or less distorted parts of the economy over time is very different from the misallocation literature's notion of allocative efficiency measured as the distance to the Pareto-efficient frontier."

[5] Indeed, Baqaee and Farhi (2020, p.107) "provide an analytical formula for the social cost of distortions, generalizing misallocation formulas like those of Hsieh and Klenow (2009) to economies with arbitrary input-output network linkages, numbers of factors, microeconomic elasticities of substitution, and distributions of distorting wedges."

The remainder of the paper proceeds as follows. Section 2 presents the model, while the model quantification is explained in Section 3. Section 4 discusses data details and empirical results for the global frontier, proxied by the US. Section 5 presents similar assessments but for the economic performance around the world. Section 6 briefly comments on two model extensions. Finally, Section 7 concludes.

2 Model

We present two versions of the model. In the first version, as usual in oligopoly models, we assume there is a discrete number of firms. A shortcoming of this version is that some key results are only *approximately* valid, as infinitesimal adjustments in the number of firms are not possible. In the second version, we consider a continuum of firms, when the standard practice is to assume these null-measure firms ignore the impacts of their decisions on aggregate outcomes, even though they exist. We follow a different approach and suppose firms pay at least some attention to them. Under this assumption, we obtain essentially the same results as in the discrete model, but holding *exactly*.

In this section, we emphasize the economic content of the model. For step-by-step derivation and formal proofs, please refer to Appendix A for the discrete version and Appendix B for the continuous one.

2.1 Discrete number of firms

We refrain from exploring households' behavior, as it is essentially irrelevant to our results; all we need is to assume more consumption is always preferred to less. Consequently, the model focuses solely on the firms' side, where misallocation originates. Additionally, since firms' decisions are static in our model, we suppress the time subscript for notational simplicity.

Environment and technology. In a closed economy, N potential entrant firms produce a single good. The price elasticity of demand for this good is strictly negative, with its absolute value denoted by η, where $1 < \eta < \infty$. One may consider there are several different goods, each produced within distinct sectors, but which can be represented by a single-sector (or single-good) economy. In this interpretation, explored in Appendix C, our model would apply to this representative sector, with η being equal to the elasticity of substitution across sectors' goods.

Since firms' goods are homogeneous, the aggregate output Y is

$$Y \equiv \sum_{i=1}^{N} Y_i, \tag{1}$$

being Y_i the production of firm i, which is given by the Cobb-Douglas function

$$Y_i = A_i K_i^{\alpha} H_i^{1-\alpha}, \tag{2}$$

where $K_i \geq 0$ is the stock of physical capital, $H_i \geq 0$ is the stock of human capital, and $A_i > 0$ is a productivity parameter, all for firm i, while $\alpha \in (0,1)$. In the following, let $\underline{\underline{A}} \equiv \min_i\{A_i\}$ and $\overline{A} \equiv \max_i\{A_i\}$ be the technology frontier of this economy, with $0 < \underline{\underline{A}} < \overline{A} < +\infty$. Denote the empirical probability of A by $g(A)$ and the corresponding empirical cumulative distribution function by $G(A) = \sum_{a<A} g(a)$.

Market competition and optimal decision. Firms engage in Cournot competition, meaning each firm chooses its output taking as given the output chosen by the other firms in the economy, as well as the wage $w > 0$ and the rental cost of physical capital $r > 0$. There are no fixed costs. Formally, each firm $i \in \{1, 2, ..., N\}$ solves the profit maximization problem

$$\begin{aligned} \max_{Y_i} \quad & (p - MC_i)\, Y_i \\ \text{s.t.} \quad & p = p(Y), Y_j \geq 0 \; \forall j \in \{1, 2, ..., N\} \setminus \{i\} \end{aligned}, \tag{3}$$

where p is the price of the good and $MC_i = \left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{w}{1-\alpha}\right)^{1-\alpha} \frac{1}{A_i}$ is the Cobb-Douglas marginal cost of firm i. The price is given by the inverse demand function $p(Y)$, with $-\left(\frac{\partial p}{\partial Y}\frac{Y}{p}\right)^{-1} \equiv \eta$. The First-Order Condition (FOC) of this optimization problem yields

$$p = MC_i \frac{\eta_i}{\eta_i - 1}, \tag{4}$$

where $\eta_i = \frac{\eta}{s_i} > 1$ is the price elasticity of demand faced by a firm with market share $s_i \equiv Y_i/Y$. Thus, the markup of firm i is $\mu_i \equiv \frac{\eta_i}{\eta_i - 1} = \left(1 - \frac{s_i}{\eta}\right)^{-1}$. Since the Second-Order Condition (SOC) for profit maximization also holds under $1 < \eta < \infty$, Equation (4) represents firm i optimal decision as long as $\mu_i \geq 1 \leftrightarrow s_i \geq 0$, that is, for every active firm i.

Equilibrium allocation. Using Equation (4) for any active firms i and j, and considering the goods are homogeneous with equalized prices across firms, it can be shown that

$$s_i - s_j = \left(1 - \frac{MC_i}{MC_j}\right)(\eta - s_j) = \left(1 - \frac{A_j}{A_i}\right)(\eta - s_j). \tag{5}$$

This implies a more productive firm will have a higher market share, as $A_j > A_i$ leads to $s_j > s_i$, given that $\eta > 1 \geq s_j$. Moreover, if A_i is sufficiently close to zero, $s_i < 0$, indicating firms with very low productivity cannot be active.[6] Thus, we must consider an entry stage to determine the set of active firms in equilibrium, even in the absence of fixed costs.

[6]To see that, rewrite (5) as $s_i = \eta - (\eta - s_j)\frac{A_j}{A_i} \leq \eta - (\eta - 1)\frac{A_j}{A_i}$, since $s_j \leq 1$, and use the fact that $\eta > 1$.

We seek an equilibrium where (i) each active firm i has non-negative profits (or, equivalently, $\mu_i \geq 1$) and (ii) non-active firms would incur strictly negative profits if they entered the market. However, this equilibrium is typically not unique (Atkeson and Burstein 2008, Edmond et al. 2015, De Loecker et al. 2021). To avoid multiple equilibria, we discard any equilibrium in which a non-active firm has a lower marginal cost than an active firm. Consequently, in equilibrium, firm i is active if and only if $A_i \geq \underline{A}$, where $\underline{A} \in \{A_1, A_2, ..., A_n\}$ is the productivity cutoff for active firms.[7] This refined equilibrium is unique and can be identified through various procedures.[8] For instance, one can apply either of the two algorithms used by Atkeson and Burstein (2008), Edmond et al. (2015), and De Loecker et al. (2021): (i) sequentially add firms in descending order of productivity, or (ii) start with all firms in the market and iteratively drop the one with the lowest negative profit.

Moreover, we consider free entry among low-productivity firms, supposing inefficient technologies are common knowledge. Specifically, we assume there is a large number of firms in any neighborhood of null productivity, meaning $\underline{\underline{A}} \to 0$ and $N \to \infty$. In such circumstances, an infinite number of low-productivity firms remain inactive, as indicated by Equation (5). This implies the profit of the marginal active firm—the firm with productivity $\underline{A}$—should be relatively low, as higher profits would incentivize new entrants. In this context, we assume the marginal firm's profit is approximately zero. Formally, for $\underline{A} = A_j$, $(p - MC_j)\, Y_j = MC_j\, (\mu_j - 1)\, Y_j \approx 0$, which holds if and only if $s_j \approx 0$, as $\mu_j = (1 - s_j/\eta)^{-1}$. Consequently, our free entry assumption implies the market share of the marginal active firm should remain close to zero, regardless of parameter values. This feature is crucial for our empirical strategy and stems from the model's strong assumptions, notably the free entry condition in a setting without fixed costs—feasible only because of the homogeneity of firms' goods. In contrast, with heterogeneous goods, firm selection requires fixed costs, leading the marginal firm to produce positive quantities in equilibrium, even under free entry. In such cases, the marginal firm's output depends on model parameters (e.g., demand elasticity and fixed costs), as identified from the near-zero-profit condition.

Evaluating Equation (5) for $A_j = \underline{A}$ and thus $s_j \approx 0$, it is easy to see that, for $A_i \geq \underline{A}$,

$$s(A_i) \equiv s_i \approx \eta\, (1 - \underline{A}/A_i)\,. \tag{6}$$

Hence, more productive firms have higher market shares and profits, while the marginal active firm, with productivity $\underline{A}$, essentially produces no output and earns no profit. By summing

[7]Strictly speaking, the criterion is firm i is active if $A_i > \underline{A}$ and only if $A_i \geq \underline{A}$, as there may be cases where not all firms with productivity $\underline{A}$ are active. For simplicity, we rule out these scenarios, resulting in only a minor loss of generality. After all, in such cases, we can slightly lower the productivity of the inactive firms with productivity $\underline{A}$, producing a nearly identical model that satisfies the stronger condition presented in the main text.

[8]To demonstrate uniqueness, assume by contradiction two procedures yield distinct sets of active firms in the refined equilibrium. Thus, necessarily, one procedure (say, the first) must exclude some lower-productivity active firms identified by the other (the second). Since the set of active firms from the first method forms an equilibrium, including these excluded firms would lead to negative profits for at least one active firm, contradicting the results of the second method.

Equation (5) over all active firms and using $s(\underline{A}) \approx 0$, we can infer the demand elasticity η from the empirical distribution of productivity among active firms:

$$\eta \approx \frac{1}{N_a \left[1 - \mathrm{E}_a \left(\underline{A}/A\right)\right]}, \tag{7}$$

where $\mathrm{E}_a \left(h(A)\right) \equiv \mathrm{E}\left(h(A)|A \geq \underline{A}\right) = \sum_{A \geq \underline{A}} h(A) \frac{g(A)}{1-G(\underline{A})}$ is the expected value of a function h over active firms under the empirical distribution, and $N_a \equiv N\left(1 - G(\underline{A})\right)$ is the number of active firms. Note Equations (6) and (7) together imply the market share of any active firm depends exclusively on the empirical productivity distribution and the cutoff for active firms. As will become clear in the next steps, this is key to measuring the misallocation of the economy without knowledge of other deep parameters.

Aggregate productivity and misallocation. From Equations (1) and (2), note

$$Y = \sum_{i=1}^{N} A_i K_i^{\alpha} H_i^{1-\alpha} = \overline{A}\Omega K^{\alpha} H^{1-\alpha}, \tag{8}$$

where $K \equiv \sum_{i=1}^{N} K_i$, $H \equiv \sum_{i=1}^{N} H_i$, and $\Omega \equiv \sum_{i=1}^{N} \theta_{Ki}^{\alpha}\theta_{Hi}^{1-\alpha}\left(A_i/\overline{A}\right)$, with $\theta_{Ki} \equiv \frac{K_i}{K}$ and $\theta_{Hi} \equiv \frac{H_i}{H}$. As firms produce homogeneous goods, the optimal allocation entails assigning all inputs to the most productive firm, when $\Omega = 1$ and $Y = \overline{A}K^{\alpha}H^{1-\alpha}$. Therefore, $\Omega \in (0, 1]$ represents the wedge between aggregate TFP, $\overline{A}\Omega$, and its optimal level, $\overline{A}$, being our measure of allocative efficiency.

Since each firm uses its inputs optimally, taking the same inputs' rental prices as given, every active firm i chooses the same physical-to-human capital ratio, and thus $\theta_{Ki} = \theta_{Hi} \equiv \theta_i$. As a result,

$$s(A_i) = \frac{Y_i}{Y} = \frac{A_i K_i^{\alpha} H_i^{1-\alpha}}{\overline{A}\Omega K^{\alpha} H^{1-\alpha}} = \frac{A_i \theta_i}{\overline{A}\Omega} \to \theta_i = \overline{A}\Omega \frac{s(A_i)}{A_i} \tag{9}$$

$$\overline{A}\Omega = \frac{1}{\sum_{i=1}^{N} \frac{s(A_i)}{A_i}}, \tag{10}$$

where we use (2) and (8) to get (9), while (10) is obtained by summing (9) over all firms. Therefore, similarly to Edmond et al. (2015) and Edmond et al. (2022), aggregate productivity is a quantity-weighted harmonic mean of firms' productivity.

Finally, by plugging Equations (6) and (7) into (10), we get

$$\Omega \approx \frac{\mathrm{E}_a \left[(\underline{A}/\overline{A})(1 - \underline{A}/A)\right]}{\mathrm{E}_a \left[(\underline{A}/A)(1 - \underline{A}/A)\right]}. \tag{11}$$

Intuitively, as in other oligopolistic models, the aggregate misallocation wedge depends on the dispersion of markups, which are directly related to firms' market shares. In our model, since market shares are determined solely by the productivity distribution, we can directly measure the allocative efficiency of the economy, Ω, provided that the productivity distribution, including its support, $\underline{A}$ and $\overline{A}$, is known. Using Equation (11), we demonstrate in Appendix A.4 that Ω improves as $\underline{A}$ increases due to the exit of less productive firms, suggesting higher productivity dispersion among active firms is associated with worse allocative efficiency, as expected. In particular, $\Omega \to 1$ as $\underline{A} \to \overline{A}$, which is intuitive since any allocation of resources is optimal if there is no productivity dispersion.[9] In this limit case, the market would resemble perfect competition, as all active firms would have unitary markups and zero profits, with $N_a \to \infty$. Interestingly, even though there are infinitely many potential entrants, the number of active firms becomes infinite only in this particular scenario. After all, from Equation (7), $N_a \to \infty$ if and only if $\mathrm{E}_a\left(\underline{A}/A\right) \to 1$, which can occur only if there is an infinite number of firms with productivity around $\overline{A}$.

Average markup. Using this model, we can also compute the cost-weighted average of firm-level markups $\mu \equiv \sum_{i=1}^{N}\left(\frac{H_i w + K_i r}{Hw + Kr}\right)\mu_i = \sum_{i=1}^{N}\theta_i\mu_i$ through

$$\mu \approx \frac{\overline{A}\Omega}{\underline{A}}. \tag{12}$$

As we discuss in Section 3, this expression is crucial for our empirical strategy.

2.2 Continuum of firms

With a discrete number of firms, some equations such as (11) are only *approximately* valid. This happens because $s(\underline{A}) > 0$ could occur in equilibrium even with free entry since the fact that all active firms are making strictly positive profits does not necessarily mean it would be profitable for a non-active (discrete) firm to enter the market. Thus, in the discrete case, one can only argue the profit of the least productive active firm is, in some sense, low—something we incorporate when using $s(\underline{A}) \approx 0$. To get equations that are *exactly* valid, we need to work with a continuum of firms. However, under such circumstances, the standard practice is to assume these null-measure firms ignore the impacts of their decisions on aggregate outcomes even though they exist. For instance, in models of monopolistic competition, each intermediate firm ignores its impact on aggregate output, but it exists as the optimal decision of final good producers relies on the marginal effects on aggregate output of employing more intermediate goods.[10] Similarly, in oligopoly models based on Atkeson and Burstein (2008), intermediate firms consider the impacts of their decisions in the outcomes of their sectors, in which a discrete number of firms compete, but overlook the impacts on aggregate variables over the continuum of sectors, which are considered only by final good producers.

[9] This result relies solely on the assumption of homogeneous goods and does not depend on the Cournot model. Please refer to Appendix A.4 for a formal proof.

[10] For the standard CES aggregator $Y = \left(\int_0^1 Y_j^{\frac{\sigma-1}{\sigma}}dj\right)^{\frac{\sigma}{\sigma-1}}$, $\partial Y/\partial Y_i = \left(\int_0^1 Y_j^{\frac{\sigma-1}{\sigma}}dj\right)^{\frac{1}{\sigma-1}} Y_i^{-\frac{1}{\sigma}} > 0$, with $\partial Y/\partial Y_i = 1$ for homogeneous goods ($\sigma \to +\infty$).

We follow a different approach, assuming each firm $i \in [0, N]$ considers $\partial Y/\partial Y_i = q \in (0, 1]$. Hence, firms pay at least some attention to their impacts on aggregate outcomes. They may be fully aware of them ($q = 1$) but cannot simply ignore them ($q = 0$). We essentially keep all other assumptions from the discrete model, with only minor adjustments. For instance, we again consider the unique equilibrium in which a firm i is active if and only if $A_i \geq \underline{A}$ for some productivity cutoff $\underline{A}$.[11] Moreover, to impose inefficient technologies are common knowledge, we assume the mass of low-productivity firms is finite but sufficiently large to the extent that not all firms can be active simultaneously. Finally, it is worth noting $\eta > 1$ no longer guarantees the SOC for firms' profit maximization, but it holds if $\eta > 0$ or if $q \in (0, 1]$ is low.[12] In the following, we simply assume these parameters satisfy that condition.

Under such conditions, all equations presented in Section 2.1 for the discrete model would still be valid, with (6), (7), (11), and (12) holding *exactly*, if one replaces (i) η by η/q and (ii) sums by integrals (e.g., $Y \equiv \int_0^N Y_i di$ and $\mathrm{E}_a\left(h(A)\right) = \int_{\underline{A}}^{\overline{A}} h(A) \frac{g(A)}{1-G(\underline{A})} dA$, where g is now a density function). These equations hold exactly because the profit of the least productive active firm is now precisely zero, due to the possibility of infinitesimal firm entries.[13] Since the two models have essentially the same equations, they have similar properties. In particular, it is easy to show that (i) $\Omega \in (0, 1]$, (ii) Ω strictly increases with $\underline{A}$, and (iii) $\Omega \to 1$ as $\underline{A} \to \overline{A}$.[14] In short, the continuous model is essentially an exact version of the discrete one.

3 Quantification strategy and baseline parameters

In this section, we outline the calibration strategy. Our primary empirical objective is to compute allocative efficiency Ω, which solely requires the distribution of active firms' productivity (Equation (11)). Thus, we aim to use real-world data to pin down the distributional parameters. We propose a calibration procedure in which, given a distributional shape, we seek $\underline{A}$ and $\overline{A}$ by matching (i) aggregate TFP $\overline{A}\Omega$ and (ii) cost-weighted average of firm-level markups μ. Owing to the simplicity of our model, we establish necessary and sufficient conditions for this procedure to work properly, achieving an exact match of both target moments. Furthermore, we elucidate the close connection between Ω and μ. Before presenting this calibration strategy, we discuss our distributional assumption and the computation of those moments using real-world data. Afterward, we present our baseline choices for the parameters, including the one that determines the shape of firms' productivity distribution.

[11] A situation in which not all firms with productivity $\underline{A}$ are active could not occur with a continuum of firms. As a consequence, in this case, we do not need to rule out such situations to obtain this result.

[12] For a discussion on this matter, please refer to Appendix B.6.

[13] In Appendix B.3, we provide a formal proof for why $s(\underline{A}) = 0$ holds in the continuous model equilibrium.

[14] See Appendix B.4 for a proof of such properties.

3.1 Distributional assumption

We consider a continuous distribution of firm productivity, consistent with the continuous model of Section 2.2. Consistent with much of the literature (e.g., Melitz and Redding 2015, Edmond et al. 2015, and Edmond et al. 2022), we employ a Pareto distribution with shape parameter $k \neq 0$, truncated within the range $A \in \left[\underline{A}, \overline{A}\right]$.[15] It is easy to show that $A \in \left[\underline{A}, \overline{A}\right]$ is also truncated Pareto distributed with shape parameter $k \neq 0$, with density $\tilde{g}(A) \equiv \frac{g(A)}{1-G(\underline{A})} = k\left(\frac{\underline{A}^k\overline{A}^k}{\overline{A}^k-\underline{A}^k}\right)A^{-k-1}$.[16] Note this density is strictly increasing in A for $k < -1$, constant for $k = -1$ (uniform distribution), and strictly decreasing for $k > -1$, $k \neq 0$. Under this density, we show in Proposition D.5 that

$$\mathrm{E}_a\left((\underline{A}/A)^j\right) = \begin{cases} \left(\frac{k}{k+j}\right)\left(\frac{\tilde{A}^{k+j}-1}{\tilde{A}^{k+j}-\tilde{A}^j}\right) & \text{, if } k+j \neq 0 \\ \left(\frac{k\tilde{A}^k}{\tilde{A}^k-1}\right)\ln\tilde{A} & \text{, if } k+j = 0 \end{cases} \tag{13}$$

for $k \neq 0$, $j \in \mathbb{N} \setminus \{0\}$, and $\tilde{A} \equiv \overline{A}/\underline{A} > 1$. This object will be useful in the mapping between average markup and the productivity distribution.

3.2 Computing the target moments using real-world data

We consider two target moments: (i) the aggregate TFP $\overline{A}\Omega$ and (ii) the cost-weighted average of firm-level markups μ. These moments can be easily obtained using standard macroeconomic data and the parameter α. The aggregate TFP is backed out as a residual in the aggregate production function (8): $\overline{A}\Omega = \frac{Y}{K^\alpha H^{1-\alpha}}$. The average markup is pinned down by the labor income share, as $\mu = \sum_{i=1}^N \theta_i\mu_i = \sum_{i=1}^N \theta_i p\frac{MPH_i}{w} = \sum_{i=1}^N \theta_i(1-\alpha)\frac{Y_ip}{\theta_iHw} = \frac{1-\alpha}{LS}$, where MPH_i is the marginal product of human capital, $MC_i = \frac{w}{MPH_i}$ as firms use inputs optimally, and $LS \equiv \frac{Hw}{Yp}$ is the labor share of national income. Note $\mu_i = \frac{1-\alpha}{LS_i}$, which is exactly Hall (1988) expression if one considers a Cobb-Douglas production function and human capital as a variable input. Hence, our method to gauge the average markup is closely related to the recent literature that uses Hall (1988) results to estimate firm-level markups (e.g., De Loecker and Warzynski 2012, De Loecker and Eeckhout 2018, De Loecker et al. 2020, Traina 2018, Calligaris et al. 2018, and Autor et al. 2020).[17]

[15] In the non-truncated Pareto distribution, k should be strictly greater than 0, but in its truncated version k could also be strictly negative.

[16] Consistently, Melitz and Redding (2015, p.24) argue that "a key feature of a Pareto distributed random variable is that it retains the same distribution and shape parameter k whenever it is truncated from below."

[17] Edmond et al. (2015) and Baqaee and Farhi (2020) use sales-weighted harmonic average instead of cost-weighted arithmetic average. However, these two measures are equivalent here, since $\mu_i = \frac{1-\alpha}{LS_i} \to \frac{s_i}{\mu_i} = \frac{Y_i/Y}{(1-\alpha)/LS_i} = \frac{H_iw/(Yp)}{1-\alpha}$ and thus the harmonic average markup $\left(\sum_{i=1}^N \frac{s_i}{\mu_i}\right)^{-1} = \frac{1-\alpha}{Hw/(Yp)} = \frac{1-\alpha}{LS} = \mu$. Similar results apply to the model of Edmond et al. (2022), in which these two average markups are also equal, given by an equivalent function of the aggregate labor share.

3.3 Calibration algorithm

Given our distributional assumption and the computed data moments, we calculate $\underline{A}$ and $\overline{A}$ by matching (i) aggregate TFP $\overline{A}\Omega$ and (ii) cost-weighted average markup μ, in each evaluated period. Hence, similarly to De Loecker et al. (2021), even though the model is static, time-varying results can be obtained due to period-by-period estimation. Formally, given $\overline{A}\Omega$ and μ, a solution for $\underline{A}$ and $\overline{A}$, if it exists, can be obtained from the following algorithm:

1. Given Equation (13), calculate $\tilde{A} \equiv \overline{A}/\underline{A}$ by numerically solving

$$\mu = \frac{1 - \mathrm{E}_a\left(\underline{A}/A\right)}{\mathrm{E}_a\left[(\underline{A}/A)(1-\underline{A}/A)\right]}, \tag{14}$$

   which is derived from (11) and (12) holding exactly as in the continuous model.

2. From Equation (12) holding exactly, compute $\underline{A} = \overline{A}\Omega/\mu$.

3. Given $\tilde{A} \equiv \overline{A}/\underline{A}$ and $\underline{A}$ from the previous steps, calculate $\overline{A} = \tilde{A} \times \underline{A}$.

In Appendix D, we establish necessary and sufficient conditions for this algorithm to work properly, achieving an exact match of both target moments. The challenge is to show that a unique solution exists for the first-step problem or, equivalently, that (14) implicitly defines $\tilde{A} \equiv \overline{A}/\underline{A}$ as a well-defined function of μ. Referring to Proposition D.8, it becomes evident a unique solution exists if and only if $\mu > 1$ and $k < 2/(\mu - 1)$.[18]

From Equation (14), the average markup μ identifies $\tilde{A}$, which is a measure of technology dispersion, reflecting the productivity gap between the most and least productive active firms. Consequently, estimating allocative efficiency Ω does not require TFP data, since (11) is solely a function of $\tilde{A}$ under (13). TFP data are used only to pin down $\underline{A}$ and $\overline{A}$. Therefore, to decompose TFP, we first estimate Ω using data on $\mu = \frac{1-\alpha}{LS}$, and then compute $\overline{A}$ from the observed TFP $\overline{A}\Omega$. As a result, in this model, the residual of the production function is not the TFP $\overline{A}\Omega$ itself, but rather only its technology component $\overline{A}$, which is a cleaner residual as it is free of misallocation effects.

Figure 1 plots the function Ω of μ for truncated Pareto distributions with $k = 3, 5, 9$ and a uniform distribution ($k = -1$). Several things are worth noting about it. First, Ω is strictly decreasing in $\mu = \frac{1-\alpha}{LS}$ and thus strictly increasing in LS. In particular, $\Omega \to 1^-$ as $\mu \to 1^+$. Essentially, this function converts one measure of distance from the efficient equilibrium into another. While μ represents the distance of average markup from the competitive, efficient, level $\mu = 1$, allocative efficiency Ω quantifies the deviation of aggregate TFP $\overline{A}\Omega$ from its optimal level, $\overline{A}$. Building on that interpretation, it seems intuitive that a lower average markup $\mu > 1$, indicating a less distorted economy, would be associated with enhanced allocative efficiency Ω. Second, given a time series of μ, a lower k would imply a higher and less volatile estimated Ω.

[18] For an arbitrary continuous truncated distribution, a unique solution exists if and only if $\mu \in \left(1, \lim_{\overline{A}\to+\infty} \mu\right)$ (Proposition D.4). By employing (14), it is possible to compute $\lim_{\overline{A}\to+\infty} \mu$, albeit contingent upon the distributional assumption. For example, for the Pareto case, this is done in Proposition D.7.

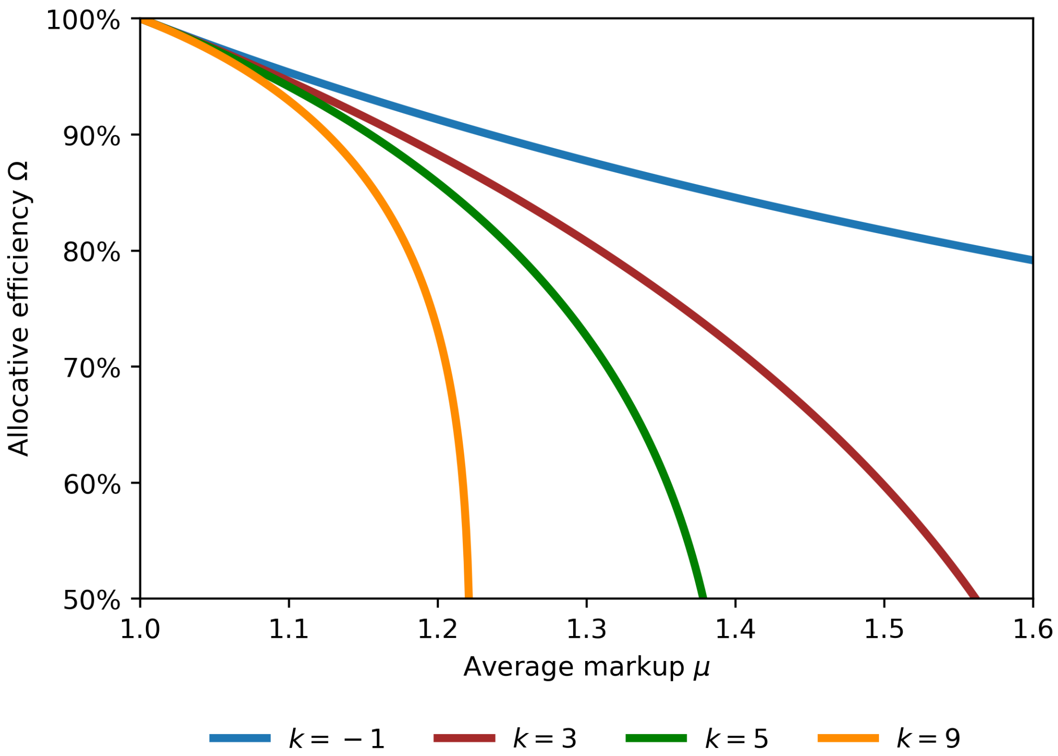

Figure 1: Allocative efficiency Ω *versus* average markup μ.

3.4 Baseline parameters

We opt to assign values to the shape parameter k because estimating it would require micro-data that is rarely available for many countries and years.[19] In our baseline calibration, we set $k = -1$, effectively assuming firm productivity is uniformly distributed.[20] This choice of k is justified for two reasons.

First, $k = -1$ yields the most conservative reasonable TFP decomposition compared to standard accounting exercises, in which all TFP variation is attributed to technology. To see that, note we obtain a TFP decomposition that is closer to standard ones by choosing a lower k, as Ω would increase towards one and become increasingly stable over time (Figure 1). Indeed, from Proposition D.9, $\Omega \to 1$ as $k \to -\infty$ and thus our model's residual $\overline{A}\Omega$ converges to the standard accounting residual $\overline{A}$ in this limit case. Given that, it would be useful to find a lower bound for k. Supposing high-productivity firms are relatively scarce, this lower bound is exactly $k = -1$. After all, as discussed in Section 3.1, the truncated Pareto density is downward sloping only for $k > -1$, $k \neq 0$, being upward sloping for $k < -1$ and constant for $k = -1$ (uniform distribution). Besides being economically reasonable, this assumption is commonly adopted, often by imposing a non-truncated Pareto distribution for firms' productivity.

Second, we show in Appendix E that, under $k = -1$, the estimates of allocative efficiency

[19]The estimation of the Pareto shape parameter is discussed in another paper (Martinez and Santos 2024), where we estimate it for Brazil using labor market concentration data.

[20]In this case, the calibration algorithm functions properly if and only if $\mu > 1$, since $k < \frac{2}{\mu - 1}$ always holds.

growth are highly robust to (i) the level of LS and (ii) the choice of α, as soon as $\mu = \frac{1-\alpha}{LS} > 1$. These are interesting features since (i) estimating α is not an easy task and (ii) gauging the level of LS is not straightforward, particularly for developing countries, given the difficulty of identifying the labor share of self-employment income (see Gollin 2002 for a discussion).

Naturally, this does not mean the choice of α is irrelevant under $k = -1$. Besides possibly changing the number of country-year observations that fulfill the model's necessary condition $\mu = \frac{1-\alpha}{LS} > 1$, α is required to pin down (i) the TFP $\overline{A}\Omega$, (ii) the (level of the) allocative efficiency Ω, and (iii) the technology frontier $\overline{A}$. As a result, the choice of α is fundamental even for decomposing the growth of TFP $\overline{A}\Omega$, as it alters the TFP itself and, consequently, the residual $\overline{A}$. In short, even for $k = -1$, one should choose α. Our baseline calibration employs the standard $\alpha = 1/3$ for all countries and years.

4 Growth in nations at the frontier

The model presented in Section 2 can be calibrated using standard macroeconomic data, as detailed in Section 3. This methodology allows us to decompose aggregate TFP into technology and allocative efficiency components, which we use to revisit key facts of economic growth. Specifically, in this section, we focus on reexamining the growth of nations at the world income frontier. Following Jones (2016), we use the United States as a proxy for this income frontier.

In Section 4.1, we discuss the data and parameters used in this exercise and validate the model by demonstrating that our misallocation estimates align with those of Edmond et al. (2022), who employ a more flexible oligopoly model. The remainder of the section is dedicated to our empirical analysis, which spans from 1954 to 2019. Section 4.2 presents the misallocation estimates, while in Section 4.3, we follow Jones (2016) in conducting a growth-accounting exercise for the US. All in all, we find misallocation can significantly impact short-run growth. For instance, between 2000 and 2007, the US experienced notable technological improvement coupled with declining allocative efficiency, suggesting the dot-com boom and advancements in IT led to productivity gains but concentrated in certain firms. Our results also indicate the technology component grows more steadily than TFP itself, at around 1% per year. Notable exceptions include the periods of 1954–1973 and 2000–2007, when technology contributed approximately 2% annually.

4.1 Data and parameters

We use annual data from the Penn World Table 10.01 (Feenstra et al. 2015) to measure the US variables. Output Y is the real GDP at constant national prices (*rgdpna* variable), and physical capital K is the capital services at constant national prices (*rkna* variable). We measure labor L as the aggregate amount of time worked, which is the product of two variables: (i) the number of persons engaged (*emp* variable) and (ii) the average annual hours worked by persons engaged (*avh* variable). Multiplying L by the human capital index (*hc* variable), based on years of schooling and returns to education, we obtain the stock of human capital H. Finally, LS is the share of labor compensation in GDP at current national prices (*labsh* variable). For the US, it is always equal to the preferred labor share measure of the Penn World Table 10.01, which assumes self-employed workers use labor and physical capital in the same proportions as the rest of the economy (Feenstra et al. 2015). The data for these variables span from 1950 to 2019, except for the physical capital stock, available only from 1954 onward.

As discussed in Section 3.4, we use the standard $\alpha = 1/3$ in the baseline calibration. Alternatively, we could calibrate it from data. Since firms use inputs optimally, α equals the cost share of capital, that is, $\alpha = \frac{Kr}{Kr+Hw}$.[21] Using factor income data for the US nonfinancial corporate sector from Barkai (2020), we estimate $\alpha = 0.31$.[22] As it is only slightly lower than the standard calibration, we maintain $\alpha = 1/3$ throughout the main body of the paper. In Appendix G, we discuss this alternative calibration procedure in detail, demonstrating that our main conclusions are robust to this lower α.

Given the labor share LS and the parameter α, we compute the average markup $\mu = \frac{1-\alpha}{LS}$. Figure 2 plots the results from 1950 to 2019, showing an increase in the average markup. This is consistent with evidence supporting higher firms' market power in the US (e.g., De Loecker et al. 2020, Baqaee and Farhi 2020, and De Loecker and Eeckhout 2018).

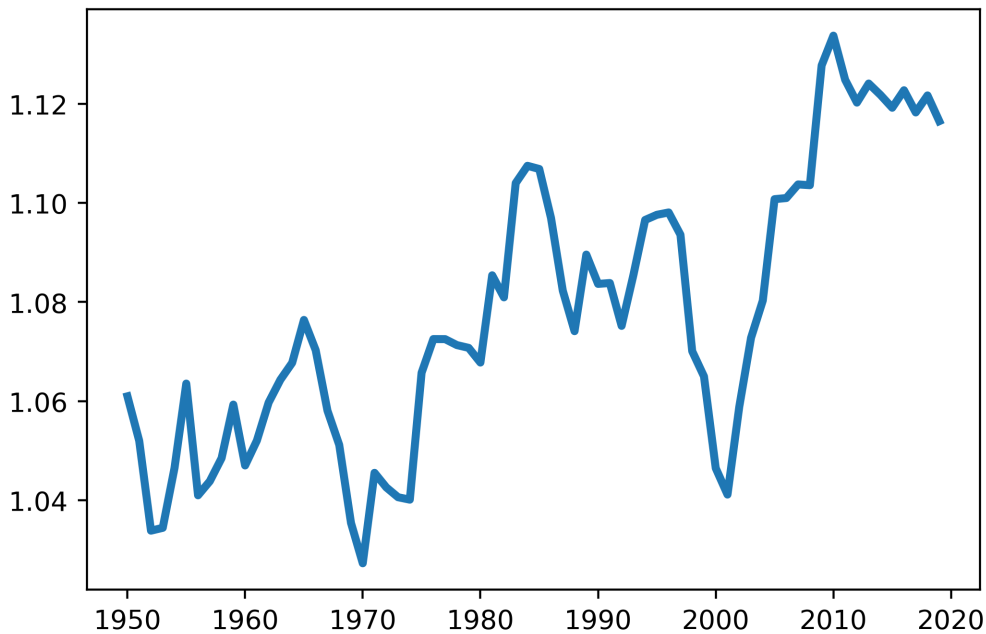

Figure 2: Average markup in the United States.

In the baseline calibration, we use $k = -1$ due to the desirable features presented in Section 3.4. However, one may be interested in evaluating the reasonableness of this assumption for the

[21]For a formal proof, please refer to Appendix A.4.

[22]Similarly, De Loecker et al. (2020) and Edmond et al. (2022) use cost data to gauge sectoral elasticity of output with respect to labor, which is equivalent to $1 - \alpha$ in our Cobb-Douglas case.

United States. Edmond et al. (2022) estimate misallocation using an oligopoly model calibrated for the US. Rather than committing to a single average markup μ, they calibrate it for μ equal to 1.05, 1.15, 1.25, or 1.35. Table 1 compares the value-added aggregate productivity losses shown in their Table 6 with our own estimates for $k = -1, 3, 5$, computed accordingly as $-(\overline{A}\Omega - \overline{A})/\overline{A} = (1 - \Omega)$. As can be seen, our estimates are closer to theirs for $k = -1$, particularly for $\mu = 1.15$ when the results are practically the same. This is a very significant result considering their model is primarily calibrated using recent data, a timeframe during which $\mu = \frac{1-\alpha}{LS}$ consistently hovers around 1.15 (Figure 2). For instance, to gauge productivity dispersion, a crucial factor in estimating misallocation, they target measures of concentration in 2012, when we find $\mu = 1.12$. This suggests $k = -1$ is indeed a reasonable assumption for the US, which serves as an important validation of our model, particularly considering the oligopoly model of Edmond et al. (2022) is more flexible, general, requiring firm-level data for calibration. In any case, as $k = -1$ is an economic lower bound for k in our model, we choose to test also the higher values shown in Table 1.

Table 1: Aggregate productivity losses, %

		Average markup μ			
		1.05	1.15	1.25	1.35
Edmond et al. (2022) estimates (Oligopoly model)		5.55	6.85	8.89	12.52
Own estimates	$k = -1$	2.41	6.75	10.54	13.90
	$k = 3$	2.60	8.43	15.31	23.58
	$k = 5$	2.70	9.62	19.87	38.77

4.2 Allocative efficiency

Figure 3 plots the estimated allocative efficiency Ω for each k. Several things are worth noting about these estimates. First, since Ω is just a transformation of μ, the estimated series are highly correlated with each other, showing similar trends.[23] In particular, all estimates suggest the allocative efficiency Ω is decreasing over time, especially after 2000, reflecting the lower labor share LS and thus the higher average markup $\mu = \frac{1-\alpha}{LS}$ in recent years (Figure 2). This is consistent with Baqaee and Farhi (2020) that finds an increase in the distance from the optimal allocation between 1997 and 2015 in the US. Second, consistent with Figure 1, Ω is higher and more stable for lower k. In any case, the estimates are relatively similar, especially considering $k = 5$ is probably very extreme given the results of Table 1.

[23]This is especially clear at the beginning of the sample, because the estimate of Ω is less sensitive to k for a lower markup μ (Figure 1), as observed in sample's initial years (Figure 2).

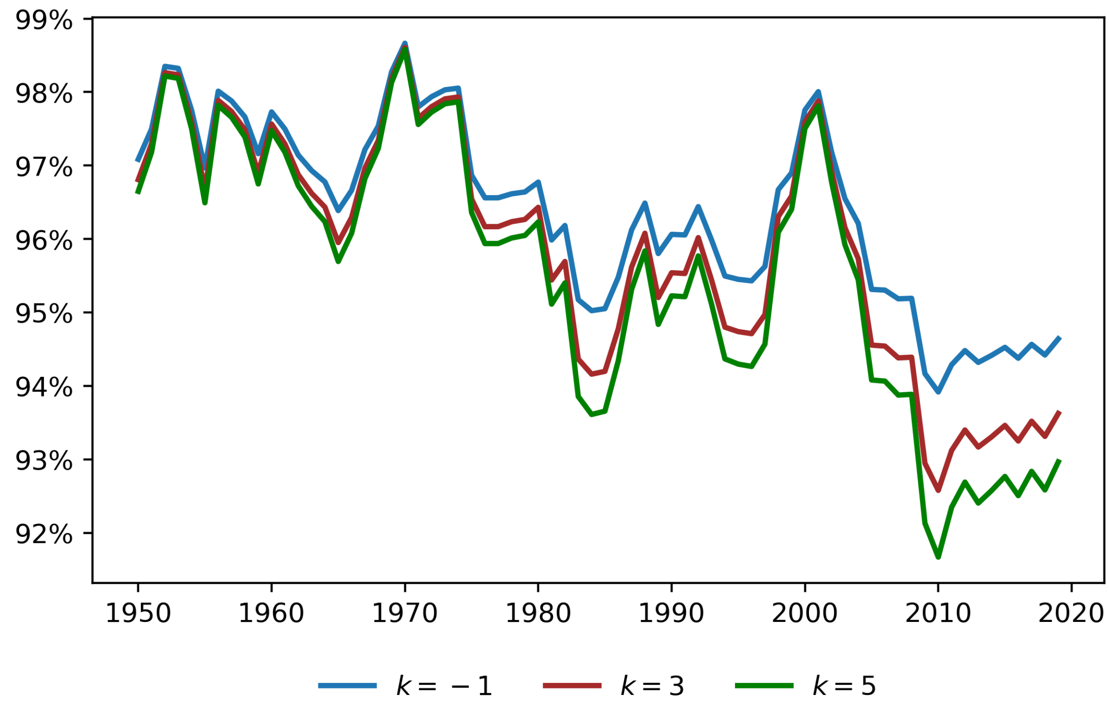

Figure 3: Allocative efficiency in the United States.

4.3 Growth accounting

To understand the sources of economic growth, one can use directly the aggregate production function (8). However, in this case, there would be no clear separation between the components since the accumulation of physical capital is typically affected by productivity (e.g., consider a standard Solow model). A better alternative followed by Jones (2016) is to use

$$\frac{Y_t}{L_t} = \left(\frac{K_t}{Y_t}\right)^{\alpha\beta} \left(\frac{H_t}{L_t}\right) \left(\overline{A}_t \Omega_t\right)^{\beta}, \tag{15}$$

which can be easily obtained from (8) if one adds time subscripts and defines $\beta \equiv \frac{1}{1-\alpha}$. In this form, output per hour Y_t/L_t grows due to the growth of (i) physical capital-output ratio through $(K_t/Y_t)^{\alpha\beta}$, (ii) human capital per hour H_t/L_t, and (iii) labor-augmenting TFP $\left(\overline{A}_t\Omega_t\right)^{\beta}$. Jones (2016) points out that the contributions from productivity and physical capital are separated in (15) in a way that they were not in (8). After all, as suggested by a neoclassical growth model, the physical capital-output ratio does not depend on TFP, at least in the long run.

Table 2 presents the growth accounting exercise based on (15). We replicate the exercise shown in Table 3 of Jones (2016) and extend it to include TFP decomposition, finding qualitatively similar results.[24] Firstly, most of the long-run growth in output per hour is attributed to TFP, although Jones (2016) reports an even higher contribution from TFP. Human capital

[24]This occurs despite the distinct data sources used. Jones (2016) uses 1948–2013 data for the US private business sector from the Bureau of Labor Statistics, while we use 1954–2019 US national data from the Penn World Table 10.01.

also plays a significant role, while the physical capital-output ratio contributes little to output growth. Additionally, we find faster output and TFP growth until 1973 as well, although our decomposition indicates the output acceleration was partially due to human capital. According to Gordon (2000), this period of outstanding TFP growth can be dated back to 1913, showcasing the impact of the great inventions of the late nineteenth and early twentieth centuries. These innovations notably encompass electricity and the internal combustion engine, but also include chemicals, movies, radio, and indoor plumbing. This boom period is followed by slower economic growth between 1973 and 1995, mainly due to a productivity slowdown. Between 1995 and 2007, economic growth accelerated again, driven by productivity growth, "coinciding with the dot-com boom and the rise in the importance of information technology" (Jones 2016, p.11). Finally, we also find lackluster economic growth from 2007 onward, in the wake of the Great Recession and among slower TFP growth. Fernald (2015) presents compelling evidence that this productivity slowdown began before the global financial crisis, tracing back to 2003, reflecting "a retreat from the exceptional, but temporary, information technology-fueled pace from the mid-1990s to early in the twenty-first century" (Fernald 2015, p.1). Similar results are found in Byrne et al. (2013).

Table 2: Growth accounting for the United States

Period	$\frac{Y}{L}$	Y/L components			Labor-aug. TFP $\left(\overline{A}\Omega\right)^{\beta}$ components					
					$k=-1$		$k=3$		$k=5$	
		$\left(\frac{K}{Y}\right)^{\alpha\beta}$	$\frac{H}{L}$	$\left(\overline{A}\Omega\right)^{\beta}$	$\overline{A}^{\beta}$	Ω^{β}	$\overline{A}^{\beta}$	Ω^{β}	$\overline{A}^{\beta}$	Ω^{β}
1954–2019	1.9	0.2	0.5	1.1	1.2	-0.1	1.2	-0.1	1.2	-0.1
1954–2013	1.9	0.2	0.6	1.2	1.2	-0.1	1.3	-0.1	1.3	-0.1
1954–1973	2.6	-0.0	1.0	1.7	1.7	0.0	1.7	0.0	1.7	0.0
1973–1990	1.3	0.3	0.5	0.5	0.7	-0.2	0.7	-0.2	0.8	-0.2
1990–1995	1.6	0.2	0.5	0.9	1.1	-0.2	1.2	-0.3	1.2	-0.3
1995–2000	2.2	0.2	0.3	1.7	1.0	0.7	0.8	0.9	0.7	1.0
2000–2007	2.2	0.4	0.3	1.5	2.0	-0.6	2.2	-0.7	2.3	-0.8
2007–2013	1.3	0.5	0.3	0.5	0.7	-0.2	0.8	-0.3	0.9	-0.4
2013–2019	1.0	-0.2	0.1	1.0	0.9	0.1	0.9	0.1	0.8	0.2

Note: Logarithmic approximation of average annual growth rates (in percent).

All in all, TFP seems to be the key variable in explaining economic growth in both the short and long run. Typically, TFP is treated as a residual, often referred to as "a measure of our ignorance" because it captures factors not directly observed. However, in our case, the residual is not the TFP $\overline{A}\Omega$ itself, but only the technology component $\overline{A}$. Allocative efficiency Ω is determined by the average markup $\mu = \frac{1-\alpha}{LS}$. Looking at Table 2, we see that long-run TFP growth is primarily driven by its technology component. Indeed, between 1954 and 2019, labor-augmenting TFP grew 1.1% per year, almost equal to the 1.2% annual growth in the technology component. Changes in allocative efficiency represented just a small drag on growth (-0.1% per year).

However, misallocation can be much more relevant for shorter periods. For instance, during 1995–2000 and 2000–2007, in the IT boom, TFP grew at similarly high rates, but for different reasons. TFP growth between 1995 and 2000 is attributable almost equally to both of its components, whereas from 2000 to 2007, it was solely driven by technology, with misallocation worsening throughout the period. These findings appear to align with sector-level TFP estimates. Fernald (2015) identifies substantial productivity gains in IT-producing industries during 1995–2000, while noting only modest improvements in IT-intensive sectors. Similarly, Gordon (2000) demonstrates that productivity enhancements from 1995 to 1999 were primarily concentrated in computer and computer-related semiconductor manufacturing. As a result, given that IT-producing sectors accounted for only 4.9% of private business value-added between 1988 and 2011 (Fernald 2015), one should not anticipate significant IT-induced aggregate productivity gains in the late 1990s. While the behavior of aggregate TFP does not align with this reasoning, our decomposition analysis does. Specifically, technology improved during 1995–2000 at essentially the same pace observed between 1973 and 1995. The TFP growth from 1995 to 2000 appears to originate from elsewhere, reflecting better diffusion of technology across firms and consequently improved resource allocation. We find significant technology improvements only after 2000, consistent with Fernald (2015) showing that TFP in the *large* IT-intensive industries surged in the early twenty-first century.[25] Nevertheless, the decline in allocative efficiency during 2000–2007 suggests these IT-induced technology improvements were not effectively disseminated across all firms.

Another example is given by the comparison between 2007–2013 and 2013–2019, when the TFP annual growth doubled, from 0.5% to 1%, but the technology component grew much more similarly across the periods due to the behavior of misallocation. Indeed, this fact seems to hold more generally, as the growth of the technology component is more stable, at around 1% per year. The main exceptions are the periods of 1954–1973 and 2000–2007, when technology contributed around 2% per year. This is consistent with the argument of Fernald (2015) that the exceptional growth during these periods is what appears unusual and warrants investigation. The subsequent slowdowns are merely the flip side of these speedups, representing the return to a normal growth pace. Additionally, among these unusual periods, it is worth noting we find more rapid technology improvements from 2000 to 2007 amidst developments in IT, even though TFP grew faster between 1954 and 1973.

[25] According to Fernald (2015), during the 1988–2011 period, these IT-intensive sectors accounted for 34.9% of private business value-added.

5 Economic performance around the world

The ultimate purpose of this paper is to revisit key facts of economic growth by using comprehensive information on misallocation across both time and space, as derived from the methodology described in Sections 2 and 3. In Section 4, we examined the growth of nations at the global income frontier, as measured in the US. Our focus now shifts to assessing economic performance worldwide, based on the data and parameters presented in Section 5.1. We approach this analysis from two perspectives. First, in Section 5.2, we conduct a development accounting exercise to understand income differences across countries. Our findings suggest misallocation plays a significant role in explaining these differences, although a substantial unexplained portion remains. Second, we reexamine the diffusion of growth across nations, actively seeking evidence of convergence not only in income but also in TFP. We find limited support for the convergence hypothesis in income and both components of TFP.

5.1 Data and parameters

The variables required by the model are sourced from the Penn World Table 10.01, which contains annual data between 1950 and 2019 for 183 countries. While development accounting solely requires variables' levels, assessing convergence involves examining the cross-country correlation between a variable's growth over a certain period and its level in the initial year. Consequently, it is essential to choose appropriate measures for both level and growth analyses, which may differ.

For level analyses, we consider two distinct data sets: one with the most accurate proxies but fewer country-year observations and another with broader coverage, albeit relying on worse measures of the variables of interest. In the data set with the greatest sample coverage, Y is the output-side real GDP at current PPPs (*cgdpo* variable), K is the capital stock at current PPPs (*cn* variable), L is the number of persons engaged (*emp* variable), H is L multiplied by the human capital index (*hc* variable), and LS is the share of labor compensation in GDP at current national prices (*labsh* variable). In the data set that uses the best proxies, Y is measured in the same way, while K is the capital services levels at current PPPs (*ck* variable). To obtain L and H, we multiply the previous proxies by the average annual hours worked by persons engaged (*avh* variable). Finally, LS is the share of labor compensation in GDP at current national prices as computed using their adjustment 2 method (*lab_sh2* variable, available at the labor detail database). This labor income share is the preferred measure of the Penn World Table 10.01, relying on the assumption that self-employed workers utilize labor and physical capital in the same proportions as the rest of the economy (Feenstra et al. 2015).

For growth analyses, we measure Y as the real GDP at constant national prices (*rgdpna* variable) and K as the capital stock at constant national prices (*rnna* variable). L and H are measured in the same way as in the data set for level analyses with the greatest sample coverage. To pin down allocative efficiency improvements, we need to measure the changes in labor share properly. However, to ensure complete coverage over the years, Feenstra et al. (2015) make some interpolations and extrapolations that distort the labor share growth as they (i) assume labor shares remain constant or (ii) linearly interpolate if there are missing years in the middle of the sample. To avoid such distortions, we choose to discard any imputed data, which are identified in their labor detail database. These imputations are done for each of the following labor share measures, which differ in the way they treat the income of self-employed workers (also known as mixed income):

1. Naïve share: $LS = \frac{\text{labor compensation of employees}}{\text{GDP at basic prices}}$.
2. Adjustment 1, mixed income: $LS = \frac{\text{labor compensation of employees}+\text{mixed income}}{\text{GDP at basic prices}}$.
3. Adjustment 2, part mixed income: $LS = \frac{\text{labor compensation of employees}}{\text{GDP at basic prices}-\text{mixed income}}$.
4. Adjustment 3, average wage: $LS = \frac{\left(\frac{\text{labor compensation of employees}}{\text{number of employees}}\right)(\text{number of employees}+\text{number of self-employed})}{\text{GDP at basic prices}}$.
5. Adjustment 4, agriculture: $LS = \frac{\text{labor compensation of employees}+\text{value added in agriculture}}{\text{GDP at basic prices}}$.

Subsequently, they apply a set of rules to select a measure for each country and year, obtaining their final estimate (*labsh* variable). Hence, the selected measure may vary from period to period, which could also distort the labor share growth. Given that, we do not use the *labsh* variable for growth analyses but choose among the above measures (without imputations). As argued for the development accounting case, adjustment 2 share (*lab_sh2* variable) is the best proxy available. However, it requires mixed income data available for only 79 countries. Consequently, we choose to measure LS using the naïve share (*comp_sh* variable, from the labor detail database) since it is available for 139 countries and seems to serve as a good proxy for the labor share growth.

This last fact is illustrated in Figure 4, which plots labor share annual growth $\Delta \ln LS$ for adjustment 2 method against each of the other four measures (all without imputations), considering only country-year observations in which all five labor share growth measures are available. We also plot the 45-degree line and the Ordinary Least Squares (OLS) linear trend, whose associated centered R^2 is shown inside each plot. Since the adjustment 1 measure also requires mixed income data, it cannot be a solution to the small sample problem of the preferred adjustment 2 share. Among the other alternatives, the naïve share has the best linear fit, with its linear trend being essentially equal to the 45-degree line. Hence, the naïve share provides the best proxy for the growth in the preferred adjustment 2 measure, suggesting it can serve as a good proxy for the labor share growth. However, the naïve share is known to underestimate the labor share as it does not allocate any income from self-employed workers to labor (Gollin 2002). As a consequence, this labor share would overestimate the average markup $\mu = \frac{1-\alpha}{LS}$, which can be relevant even for growth analyses. For instance, for $k > 0$, a country may even be dropped from the sample since the necessary condition $k < \frac{2}{\mu-1}$ could not hold. To address this issue, we multiply the naïve share (without imputation) by a country-specific constant that aligns its average with the mean of the *labsh* variable (with imputation), using the longest coincident period available. By doing that, we maintain the naïve share growth but correct its level.

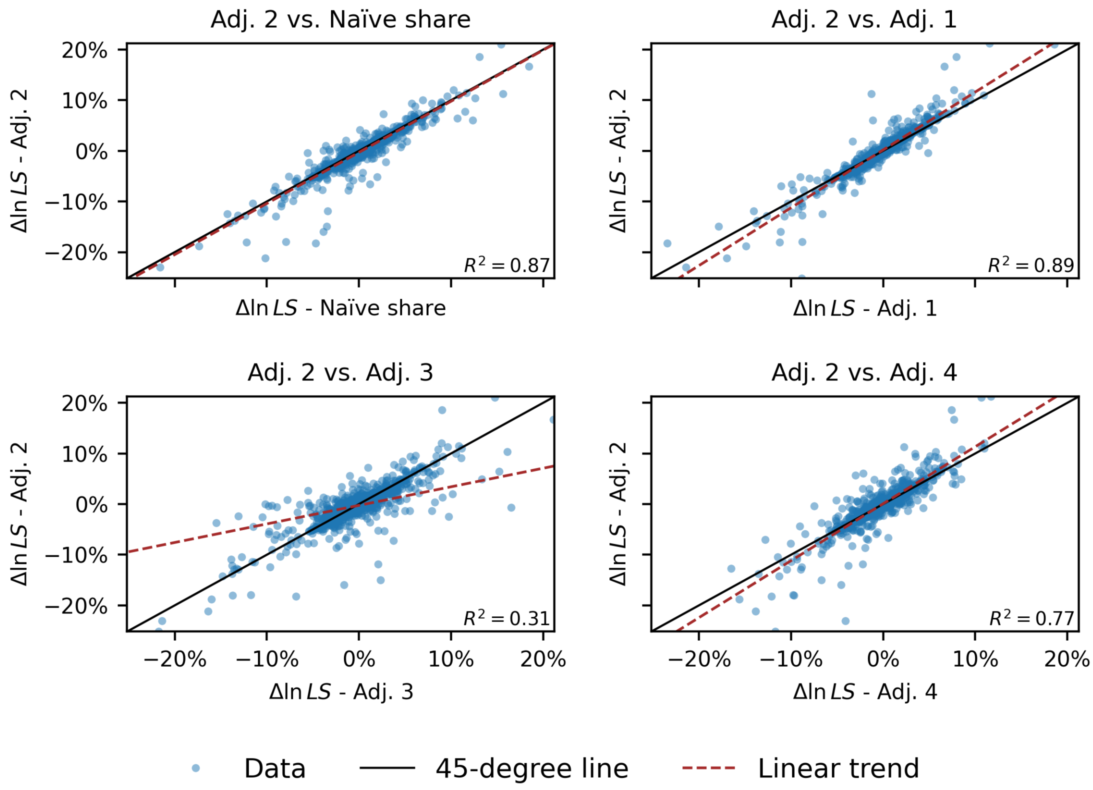

Figure 4: Comparing measures of labor share annual growth.

We constructed this data set for growth analyses focusing on maximizing sample coverage. Additionally, we considered another data set that employs the best proxies available for growth analyses.[26] However, it is impractical to evaluate convergence using the best proxies for both level and growth analyses due to the significantly reduced sample size. As a result, we present convergence results only for the largest data sets.

We finish the data description by commenting on two variables used in both level and growth analyses. First, GDP per person, which is Y divided by the population (*pop* variable from the Penn World Table 10.01). Second, our analyses are focused on non-oil countries. We classify a country as an oil producer in a given year if its oil rents account for more than 10% of GDP in that year. Oil rents as a share of GDP are sourced from the World Development Indicators provided by the World Bank.[27]

Regarding the parameters, we use the baseline values $\alpha = 1/3$ and $k = -1$ discussed in

[26] In this case, K is the capital services at constant national prices (*rkna* variable), L and H are the best proxies for level analyses, and LS is the adjustment 2 share (*lab_sh2* variable), without imputations.

[27] In cases where oil rents are unavailable for a country at a given year, we utilize data from the most recent non-missing year for that country. If such data are unavailable, we resort to data from the subsequent non-missing year. If these data are also unavailable, we assume oil rents to be zero.

Section 3.4.[28] Since $k = -1$ is an economic lower bound, we also test $k = 3$.

Figure 5 plots the number of non-oil countries for which we could decompose the TFP for each data set, k, and year between 1950 and 2019. Several facts of this graph stand out. First, as expected, the best proxies are available for fewer countries. In particular, the sample coverage of the data set with the best proxies for growth analyses is notably restricted, especially before the mid-1990s. Second, the sample coverage is smaller for $k = 3$ than for $k = -1$, since the necessary condition $k < \frac{2}{\mu - 1}$ always holds under $\mu > 1$ for $k < 0$, but not necessarily for $k > 0$.[29] Third, the sample coverage typically increases over time. The main exception occurs in growth analyses for 2019 when the labor share measures are available only for the US. This does not happen in the data sets for level analyses due to the use of extrapolated data. Fourth, this use of interpolated/extrapolated labor share data in level analyses increases its sample coverage relative to growth analyses, which is especially clear closer to the beginning of the sample.

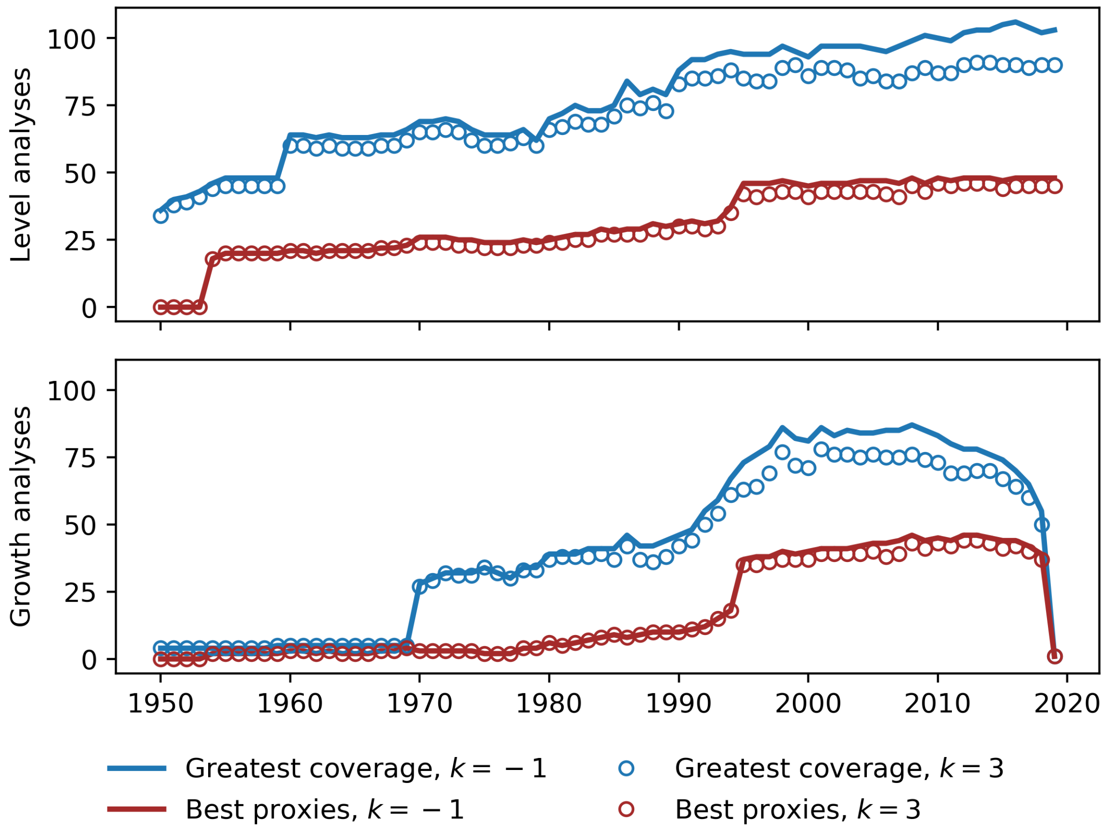

Figure 5: Number of non-oil countries whose TFP could be decomposed.

In the subsequent sections, we leverage these data, employing the TFP decomposition to reexamine key aspects of economic growth. For an initial exploration of the data sets for level

[28] Alternatively, we could have used $\alpha = 0.31$, as estimated in Section 4.1 using US data from Barkai (2020). However, since this value is not significantly different from $\alpha = 1/3$, we opt to adhere to the standard practice employed in development accounting exercises (e.g., Caselli 2005, 2016; and Jones 2016).

[29] Indeed, for $\mu > 1$ and $k = 3 > 0$, $k < \frac{2}{\mu-1} \leftrightarrow \mu < \frac{2+k}{k} \approx 1.67$.

analyses, please refer to Appendix H. This appendix features scatter plots depicting the labor-augmenting TFP and its components against GDP per unit of labor for 1965, 1975, 1985, 1995, 2005, and 2015. Appendix I provides a similar exploration of the data sets for growth analyses, but, in this case, we plot each variable's five-year annual average growth instead of their levels.

5.2 Development accounting

Growth accounting aims at identifying the proximate determinants of economic growth. It is commonly utilized to evaluate a country's performance over time, as we do for the US in Section 4.3. Development accounting applies the same logic but to explain income differences across countries at a specific point in time, typically a year. As Jones (2016) points out, the physical capital-output ratio is remarkably stable across countries. Human capital varies more as poorer countries usually present lower levels of educational attainment, but still modestly so in light of the substantial cross-country income differences. As a consequence of these two facts, wealthier nations exhibit significantly higher productivity, leading to the usual conclusion that income differences are predominantly attributed to TFP (e.g., Caselli 2005, 2016; and Jones 2016). This prompts the question: can misallocation offer insights into cross-country income differences?

Following the discussion of Section 4.3, our development accounting exercise relies on production function (15) instead of (8), since, in that case, the contributions from productivity and physical capital are better separated than in this one.[30] From (15),

$$\mathrm{Var}(y) = \mathrm{Var}(e) + \mathrm{Var}(u) + 2\mathrm{Cov}(e,u), \tag{16}$$

where $\mathrm{Var}(\cdot)$ and $\mathrm{Cov}(\cdot)$ are, respectively, sample variance and covariance across countries at a given year, while $y \equiv \ln(Y/L)$ is the natural logarithm of the income per unit of labor, e represents the explained portion of y, and u denotes the unexplained part of y. We initially consider the standard case where TFP is a residual, such that $e \equiv \ln\left[(K/Y)^{\frac{\alpha}{1-\alpha}}(H/L)\right]$ and $u \equiv \ln\left[\left(\overline{A}\Omega\right)^{\frac{1}{1-\alpha}}\right]$. Based on this variance decomposition, we can identify how successful the factors of production are in explaining cross-country income differences. We use two standard measures from the development accounting literature (see, e.g., Klenow and Rodríguez-Clare 1997, and Caselli 2005):

$$success_1 \equiv \frac{\mathrm{Var}(e)}{\mathrm{Var}(y)} \tag{17}$$

$$success_2 \equiv \frac{\mathrm{Var}(e) + \mathrm{Cov}(e,u)}{\mathrm{Var}(y)} = \frac{\mathrm{Cov}(e,y)}{\mathrm{Var}(y)}. \tag{18}$$

Undoubtedly, the variance term associated with the factors of production should be incorpo-

[30] However, here one should interpret the subscript t in (15) as denoting country, not time.

rated into the numerator of such success measures. However, the appropriate treatment of the covariance term is less clear and distinguishes these two metrics. In the first measure, the entire covariance term is assigned to TFP, whereas the second measure distributes this term equally between factors of production and TFP. Note this second measure can also be computed as the OLS slope coefficient from regressing the factors term e on the output per unit of labor y. Analogous reasoning applies to its complement, the TFP share. Hence, as Klenow and Rodríguez-Clare (1997, p.80) point out, this variance decomposition amounts to asking: when we see 1% higher output per labor in one country, how much higher we expect factors and TFP terms to be?

Alternatively, we calculate these two success measures with allocative efficiency included in the explained portion of y, consistent with our oligopoly model in which the residual is just the technology component $\overline{A}$ rather than the entire TFP $\overline{A}\Omega$. Formally, we again employ (17) and (18), but, in this case, using $e \equiv \ln\left[(K/Y)^{\frac{\alpha}{1-\alpha}}(H/L)\,\Omega^{\frac{1}{1-\alpha}}\right]$ and $u \equiv \ln\left[\overline{A}^{\frac{1}{1-\alpha}}\right]$. For both measures, high values indicate e successfully explains cross-country income differences, while low values denote failure as u is the key variable behind the observed results. Consequently, by computing these metrics both with and without allocative efficiency Ω included in the explained portion e, we can check if misallocation enhances our understanding of cross-country income differences. If the measures are higher when Ω is included in e, we conclude it enhances our understanding; if they are not, we conclude it does not.

The results are displayed in Figure 6. We use all non-oil countries available for each data set for level analyses, k, and year. Therefore, as previously seen in Figure 5, the evaluated countries may change over time in a given plot and across plots for a given year, requiring caution when performing such comparisons. Our primary focus is comparing the measures within a given data set, k and year, when the same sample of countries is used. As can be seen, both measures of success increase when misallocation is included in the explained portion, across all the evaluated years. This is evident in Figure 7, which plots the difference between the metrics with and without misallocation. The gains are more pronounced for the first measure. This means the inclusion of misallocation elevates the variance of the explained portion, thereby enhancing the first metric. However, the covariance between explained and unexplained components diminishes, resulting in smaller improvements in the second measure. Or, equivalently, both the variance of the explained portion and its covariance with income per unit of labor increase, but the former sees a greater rise. Note this higher covariance implies a positive relationship between allocative efficiency and income, which is illustrated in the scatter plots of Appendix H (Figures H.9, H.10, H.11, and H.12).

It is also noteworthy that the gains are more modest for $k = -1$, typically below 10pp, but it becomes more relevant as k increases. Moreover, we find stronger enhancements when using the best proxies available, with the gains typically doubling in this smaller data set. However, this last result may be reflecting differences in countries' sample rather than the measures themselves, as these increased gains essentially disappear when we consider the same sample

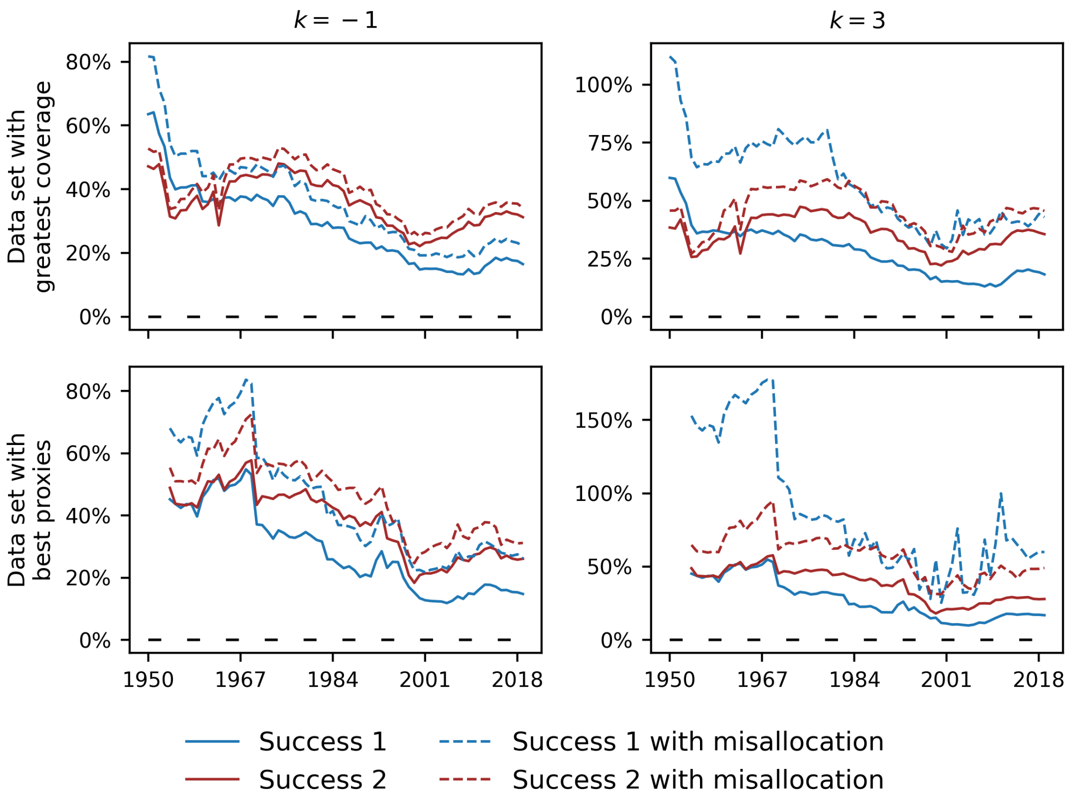

Figure 6: Measures of success, all non-oil countries available.

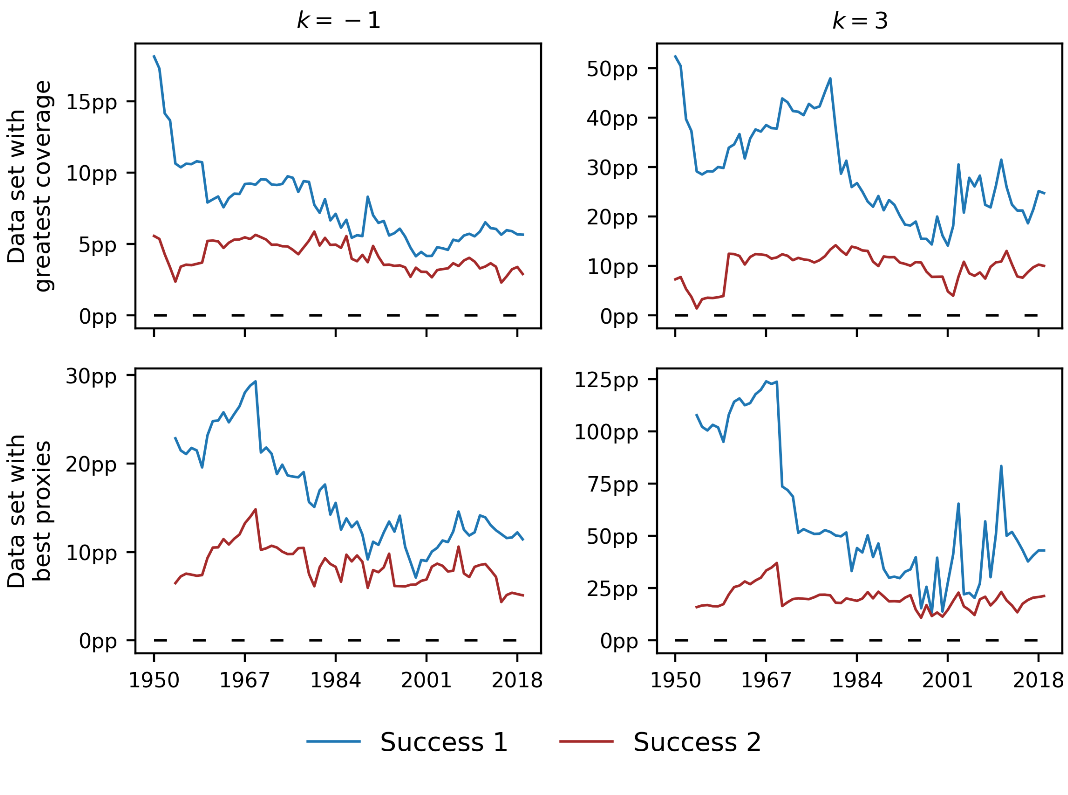

Figure 7: Success gains due to misallocation, all non-oil countries available.

of countries in each data set (Figures J.1 and J.2 in Appendix J).[31] In contrast, the higher gains for $k = 3$ are not attributable to sample issues; they instead reflect the characteristics of the model shown in Figure 1, with a higher k leading to allocative efficiency estimates that vary more across countries.

In summary, these metrics consistently indicate misallocation plays a significant role in explaining cross-country income differences, particularly for higher values of k. Despite its significance, the measures of success remain relatively low, typically below 50%, at least for the most recent two decades, which have broader sample coverage. In other words, the unexplained portion continues to be substantial, still constituting the majority of observed variability in most cases. Caselli (2005, p.681) states that a "sentence commonly used to summarize the existing literature sounds something like 'differences in efficiency account for at least 50% of differences in per capita income"'. Our results indicate that while market-power-driven misallocation is essential to understanding cross-country income differences, it does not fundamentally alter this statement.

5.3 Convergence assessment

In his Figure 26, Jones (2016) evaluates income convergence by comparing the 1960–2011 economic growth rate with GDP per person in 1960 across countries. As these variables do not exhibit a negative correlation, Jones (2016, p.36) concludes "that a simplistic view of convergence does not hold for the world as a whole." In Figure 8, we conduct a similar exercise but for 1975–1995, 1975–2005, 1975–2015, 1985–2005, 1985–2015, and 1995–2015. GDP per person growth (level) is from the data set for growth analyses (level analyses) with the greatest coverage. We consider only countries whose TFP could be decomposed in both level and growth data sets under $k = -1$.[32] For each period, we also show OLS linear trends for non-oil countries and the complete set of countries, whose slope coefficients' estimates and statistical significance are displayed inside each plot. Consistent with Jones (2016), we do not find strong support for income convergence. Indeed, the variables seem to be negatively correlated only for 1995–2015.

In Section 5.2, we saw that TFP is key in understanding cross-country output differences, suggesting the above findings for income may reflect a lack of convergence in TFP. We investigate it in Figures 9, 10, and 11, where we perform convergence assessments for TFP $\overline{A}\Omega$, technology frontier $\overline{A}$, and allocative efficiency Ω, respectively. Similarly to the income case, the convergence hypothesis does not seem to hold for the TFP and both its components except perhaps for 1995–2015, when poorer countries grew faster among swifter technology advancements and more pronounced gains in allocative efficiency.

[31] In Appendix J, we also depict the success measures from 1990 onwards but under time-invariant countries' samples (Figures J.3 and J.4). Overall, the results exhibit consistency with those presented here.

[32] We distinguish between oil-producing and non-oil-producing countries in the plot. An oil-producing country is defined as one where oil rents account for more than 10% of GDP in *any* year within the specified period.

A more comprehensive convergence assessment is presented in Figure 12, which shows the same non-oil slope estimates displayed in the scatter plots but for each 20-year period commencing in 1970. We additionally depict the 90% confidence interval. All in all, this evidence reaffirms our earlier observations that convergence appears elusive, except possibly in the most recent decades. It is noteworthy these conclusions remain valid for $k = 3$ (Figure 13).[33]

In short, the empirical evidence does not strongly support the convergence hypothesis in either income or TFP, with the lack of convergence being evident in both TFP components. Consequently, countries do not appear to be converging over time to a common degree of allocative efficiency, indicating the level of efficiency is country-specific even in the long run. Interestingly, this suggests market-power-driven misallocation is linked, in the long run, to long-lasting country-specific factors such as institutions.

6 Model extensions

We discuss two model extensions in Appendix F. First, we go beyond the Cobb-Douglas production function (2) and show that the model's key equations are still valid for an arbitrary well-behaved production function with M factors of production, provided it exhibits (i) constant returns to scale and (ii) Hicks-neutral productivity shifter. Consequently, given data on the TFP $\overline{A}\Omega$ and the average markup μ, we can quantify the model by following precisely the empirical strategy of Section 3.3, relying on the same conditions for the existence of a solution for the calibration algorithm. The differences between these models appear only in the computation of the TFP $\overline{A}\Omega$ and the average markup μ from data since (i) the expressions used to compute these two data moments are different and (ii) the calibration of the production function using cost share data may require more than just the long-run average or median used in Section 4.1.

Second, we add firm-specific wedges as a new source of firm heterogeneity, which could be helpful, for instance, to study the impact of size-dependent policies on misallocation. More precisely, we consider firm-specific tax rate over revenue $\tau_i = \tau(A_i)$. In this case, model equations change, and consequently, the conditions for the existence of a solution for the calibration algorithm of Section 3.3 are no longer valid. We do not obtain new such conditions but briefly comment on the empirical implementation of this model when (i) the function τ is known and (ii) the function τ is known except for a parameter, but the sales-weighted average tax rate $\tilde{\tau} \equiv \sum_{i=1}^{N} s_i \tau_i$ is observed.

[33] As mentioned in Section 5.1, we could not conduct such thorough investigations using our best proxies due to their limited sample size, especially before the mid-1990s. Nonetheless, if anything, more recent data appear to corroborate these findings.

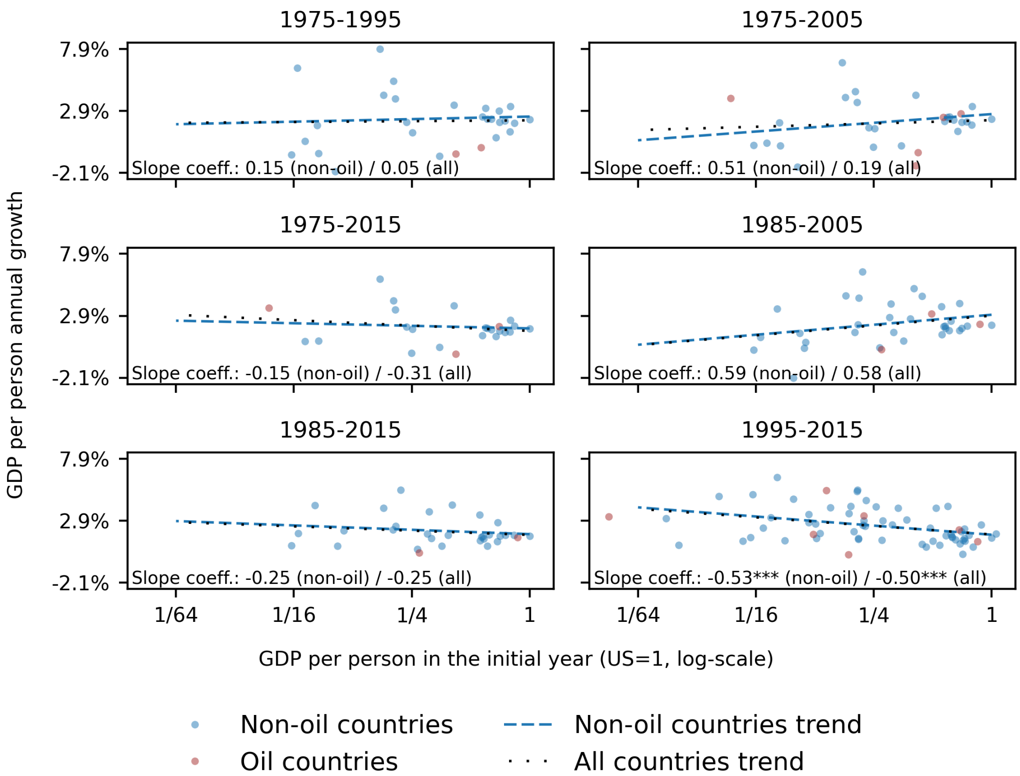

Figure 8: Assessing convergence in income – data sets with greatest coverage and $k = -1$. *Note*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

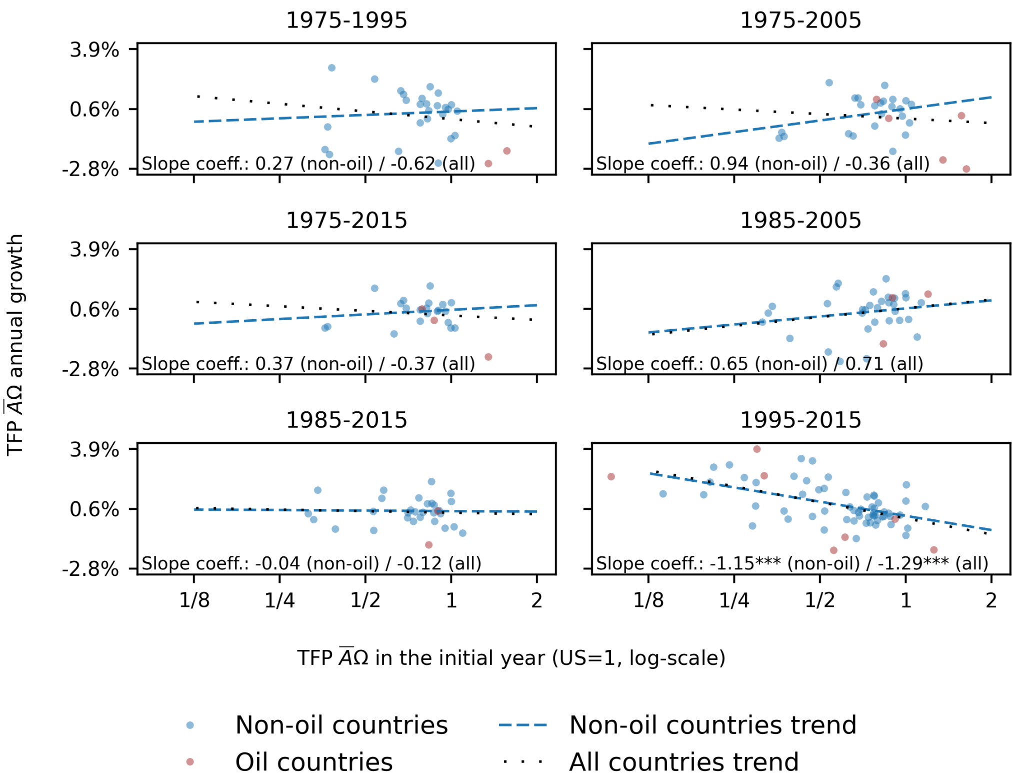

Figure 9: Assessing convergence in TFP $\overline{A}\Omega$ – data sets with greatest coverage and $k = -1$. *Note*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

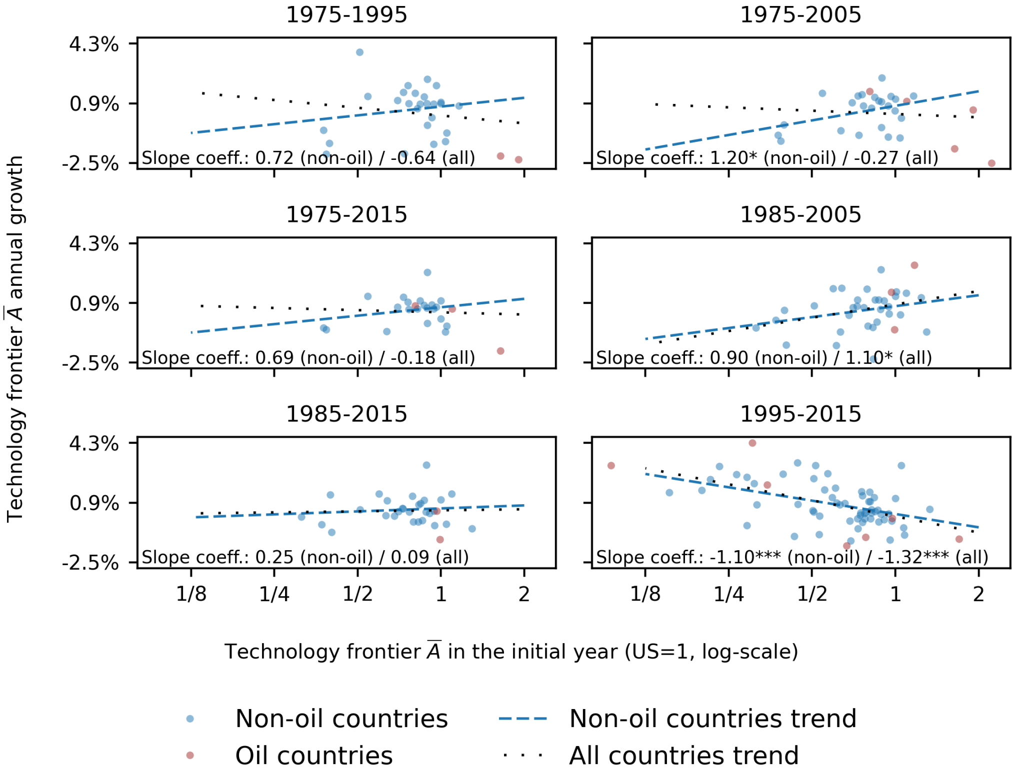

Figure 10: Assessing convergence in technology frontier $\overline{A}$ – data sets with greatest coverage and $k = -1$. *Note*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

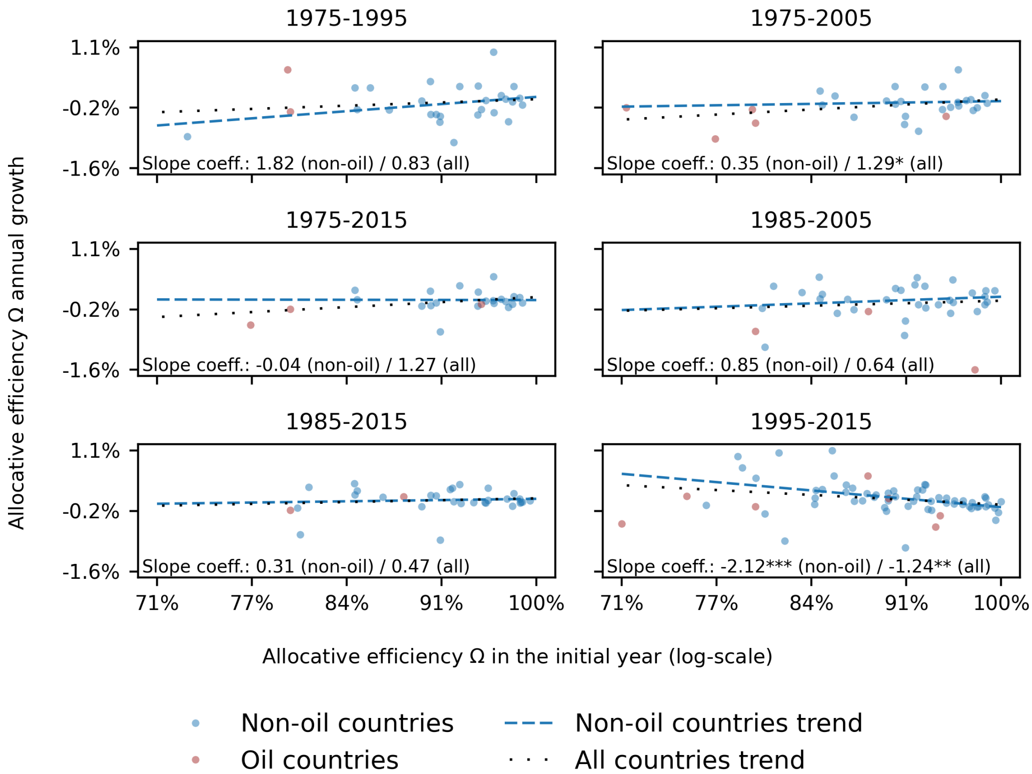

Figure 11: Assessing convergence in allocative efficiency Ω – data sets with greatest coverage and $k = -1$. *Note*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

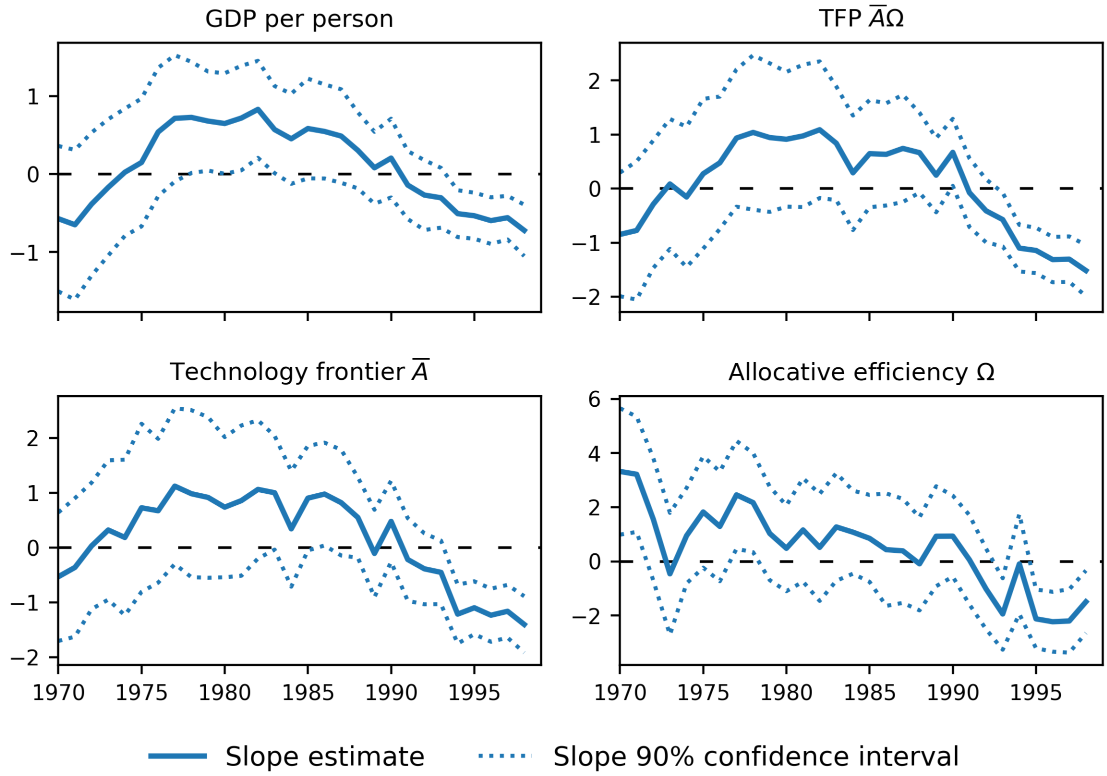

Figure 12: Slope coefficient from regressing variable's annual growth over the next 20 years (in percent) on its level (in natural logarithm) – data sets with greatest coverage and $k = -1$.

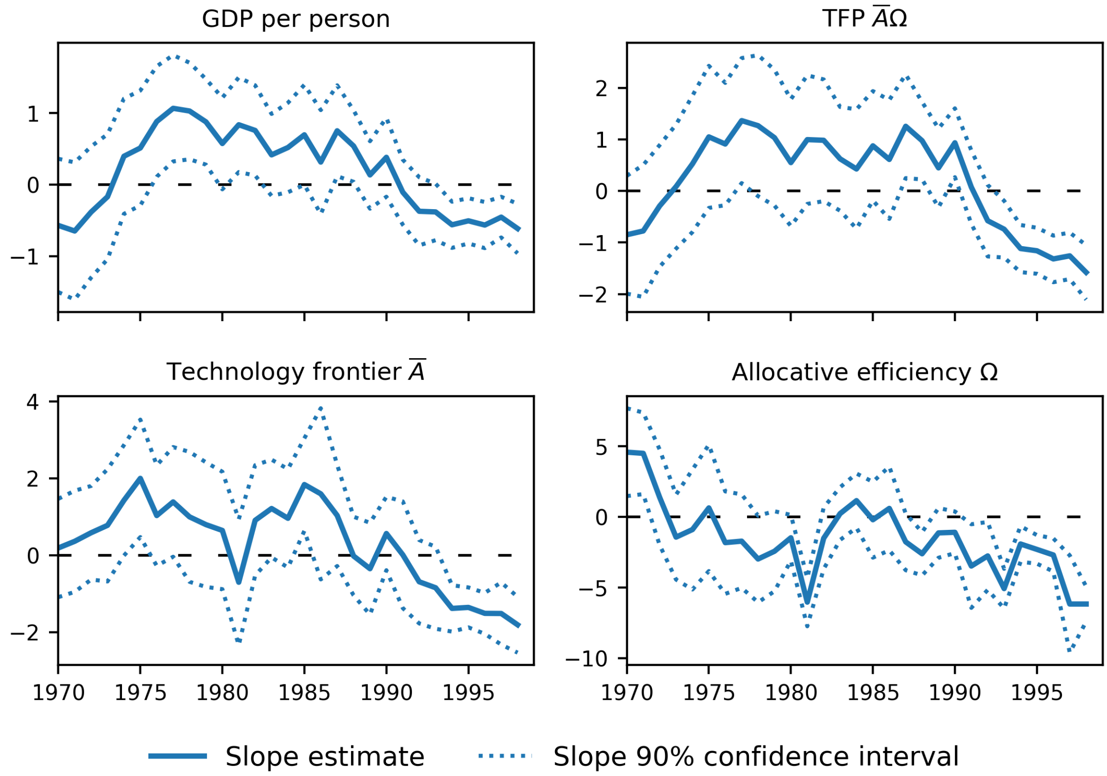

Figure 13: Slope coefficient from regressing variable's annual growth over the next 20 years (in percent) on its level (in natural logarithm) – data sets with greatest coverage and $k = 3$.

7 Conclusion

Using a Cournot model, this study decomposes TFP into technology and allocative efficiency components from 1950 to 2019 for up to a hundred countries from the Penn World Table 10.01. This decomposition enables a reexamination of key facts of economic growth. Our evaluation of the world income frontier, proxied by the US, reveals that changes in misallocation can significantly impact short-run growth. For example, during 2000–2007, the US witnessed notable technological improvement coupled with declining allocative efficiency, suggesting the dot-com boom and advancements in IT led to productivity gains but concentrated in certain firms. On a more general note, the technology component seems to grow more steadily than the TFP itself, at around 1% per year. Notable exceptions are the periods of 1954–1973 and 2000–2007, when technology contributed approximately 2% annually.

Turning to a global perspective, our analysis suggests misallocation plays a significant role in explaining cross-country income differences. Including misallocation increases our first measure of success by raising the variance of the explained portion. We also observe an increased covariance between the explained portion and the income per unit of labor when allocative efficiency is considered. This explains why our second measure of success increases as well, even though the improvements are smaller in this case. Despite its significance, a considerable unexplained portion persists, constituting the majority of observed variability in most cases. Finally, we obtain limited support for the convergence hypothesis in income and both components of TFP. Interestingly, the lack of convergence in allocative efficiency suggests market-power-driven misallocation is linked, in the long run, to long-lasting country-specific factors such as institutions.

Although the focus of this paper is normative, we would like to conclude with a brief discussion of policy implications. In our model, policies aimed at reducing market-power-driven misallocation are not necessarily desirable. For example, while forcing the most productive firms to exit the market would increase allocative efficiency, it would certainly *reduce* aggregate productivity rather than enhance it. The optimal approach in our model is to ensure the rapid diffusion of the best available technologies among firms. However, this conclusion should be approached with caution, as we treat technology as exogenous. If this process were endogenized, the result could differ. For instance, in endogenous growth models, market power or abnormal profits often serve as a reward for the risky and costly task of innovating (e.g., Romer 1990, and Aghion and Howitt 1992), implying slower technology diffusion would create stronger incentives for technological advancement. This suggests a trade-off between higher allocative efficiency with fast technology diffusion and better technological development with slower diffusion. Overall, technology diffusion should strike a balance—neither too fast nor too slow. While this issue is crucial for policy analysis, it lies beyond the scope of this paper and is left for future research. In any case, this trade-off does not appear to be the primary driver of cross-country income differences, as our empirical findings indicate that low-income countries typically experience both higher misallocation and technological backwardness.

Appendices

A Derivation of the discrete model

We refrain from exploring households' behavior, as it is essentially irrelevant to our results; all we need is to assume more consumption is always preferred to less. Consequently, the model focuses solely on the firms' side, where misallocation originates. Additionally, since firms' decisions are static in our model, we suppress the time subscript for notational simplicity.

A.1 Environment and technology

In a closed economy, N potential entrant firms produce a single good. The price elasticity of demand for this good is strictly negative, with its absolute value denoted by η, where $1 < \eta < \infty$. Since firms' goods are homogeneous, the aggregate output Y is

$$Y \equiv \sum_{i=1}^{N} Y_i, \tag{1}$$

being Y_i the production of firm i, which is given by the Cobb-Douglas function

$$Y_i = A_i K_i^{\alpha} H_i^{1-\alpha}, \tag{2}$$

where $K_i \geq 0$ is the stock of physical capital, $H_i \geq 0$ is the stock of human capital, and $A_i > 0$ is a productivity parameter, all for firm i, while $\alpha \in (0,1)$. In the following, let $\underline{\underline{A}} \equiv \min_i\{A_i\}$ and $\overline{A} \equiv \max_i\{A_i\}$ be the technology frontier of this economy, with $0 < \underline{\underline{A}} < \overline{A} < +\infty$. Denote the empirical probability of A by $g(A)$ and the corresponding empirical cumulative distribution function by $G(A) = \sum_{a<A} g(a)$.

A.2 Market competition and optimal decision

Firms engage in Cournot competition, meaning each firm chooses its output taking as given the output chosen by the other firms in the economy, as well as the wage $w > 0$ and the rental cost of physical capital $r > 0$. Formally, each firm $i \in \{1, 2, ..., N\}$ solves the profit maximization problem

$$\begin{array}{ll} \max\limits_{Y_i} & (p - MC_i)\, Y_i \\ \text{s.t.} & p = p(Y), Y_j \geq 0\, \forall j \in \{1, 2, ..., N\} \setminus \{i\} \end{array}, \tag{3}$$

where p is the price of the good and $MC_i = \left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{w}{1-\alpha}\right)^{1-\alpha} \frac{1}{A_i}$ is the Cobb-Douglas marginal cost of firm i. The price is given by the inverse demand function $p(Y)$, with $-\left(\frac{\partial p}{\partial Y}\frac{Y}{p}\right)^{-1} \equiv \eta$.

The First-Order Condition (FOC) of this optimization problem is

$$
\begin{aligned}
0 =& p\left(1+\frac{\partial p}{\partial Y_i}\frac{Y_i}{p}\right)-MC_i = p\left(1-\frac{1}{\eta_i}\right)-MC_i \\
p =& MC_i\frac{\eta_i}{\eta_i-1},
\end{aligned}
\tag{4}
$$

where $\eta_i \equiv -\left(\frac{\partial p}{\partial Y_i}\frac{Y_i}{p}\right)^{-1}$ is the absolute value of the price elasticity of demand faced by firm i. Being $s_i \equiv \frac{Y_i}{Y}$ the market share of firm i, note $\eta_i = \left(-\frac{\partial p}{\partial Y}\frac{\partial Y}{\partial Y_i}\frac{Y}{p}\frac{Y_i}{Y}\right)^{-1} = \frac{\eta}{s_i} > 1$. As a result, the markup of firm i is $\mu_i \equiv \frac{\eta_i}{\eta_i - 1} = \left(1-\frac{s_i}{\eta}\right)^{-1}$.

The Second-Order Condition (SOC) is

$$
\begin{aligned}
0 >& \frac{\partial p}{\partial Y}\frac{\partial Y}{\partial Y_i}\left(1-\frac{s_i}{\eta}\right)-\frac{p}{\eta}\left(\frac{Y-Y_i\frac{\partial Y}{\partial Y_i}}{Y^2}\right) \\
0 >& -\frac{p}{\eta Y}\left(1-\frac{s_i}{\eta}\right)-\frac{p}{\eta Y}(1-s_i) = -\frac{p}{\eta^2 Y}\left[\eta(2-s_i)-s_i\right] \\
\eta >& \frac{s_i}{2-s_i},
\end{aligned}
\tag{A.1}
$$

which is satisfied for $\eta > 1$, as $\frac{s_i}{2-s_i}$ is strictly increasing in s_i and equals 1 for $s_i = 1$.

Therefore, Equation (4) represents firm i optimal decision as long as $\mu_i \geq 1 \leftrightarrow s_i \geq 0$, that is, for every active firm i.

A.3 Equilibrium allocation

Using Equation (4) for any active firms i and j, and considering prices are equalized across firms, note

$$
\begin{aligned}
MC_i(\eta-s_i)^{-1} =& MC_j(\eta-s_j)^{-1} \\
s_i - s_j =& \left(1-\frac{MC_i}{MC_j}\right)(\eta-s_j) = \left(1-\frac{A_j}{A_i}\right)(\eta-s_j).
\end{aligned}
\tag{5}
$$

As discussed in the main text, in the unique refined equilibrium, there exists a firm with productivity $\underline{A}$ serving as the cutoff for active firms, such that firm i is active if and only if $A_i \geq \underline{A}$. Given that, rewrite (5) as $\frac{s_i}{A_j} = \frac{\eta}{A_j} - \frac{1}{A_i}(\eta - s_j)$ and sum it in j over all active firms to obtain

$$
\begin{aligned}
s_i\mathrm{E}_a(1/A) =& \eta\mathrm{E}_a(1/A) - \frac{1}{N_a A_i}(N_a\eta - 1) \\
s(A_i) \equiv s_i =& \frac{1}{N_a\mathrm{E}_a(A_i/A)} + \eta\left(1-\frac{1}{\mathrm{E}_a(A_i/A)}\right),
\end{aligned}
\tag{A.2}
$$

where $\mathrm{E}_a(h(A)) \equiv \mathrm{E}(h(A)|A \geq \underline{A}) = \sum_{A\geq\underline{A}} h(A)\frac{g(A)}{1-G(\underline{A})}$ is the expected value of a function h over active firms under the empirical distribution, and $N_a \equiv N(1-G(\underline{A}))$ is the number of active firms. As expected, these shares add to one over all active firms.

Given the free-entry assumption, we use $s(\underline{A}) \approx 0$, implying that, for $A_i \geq \underline{A}$,

$$s(A_i) \approx \eta \left(1 - \underline{A}/A_i\right) \tag{6}$$

$$\eta \approx \frac{1}{N_a \left[1 - \mathrm{E}_a\left(\underline{A}/A\right)\right]}, \tag{7}$$

where we obtain (6) by plugging $A_j = \underline{A}$ and $s_j \approx 0$ into (5), while we use $s(\underline{A}) \approx 0$ in (A.2) to get (7).

A.4 Aggregate productivity and misallocation

Given (1) and (2), note

$$Y = \sum_{i=1}^{N} A_i K_i^{\alpha} H_i^{1-\alpha} = \overline{A}\Omega K^{\alpha} H^{1-\alpha}, \tag{8}$$

where $K \equiv \sum_{i=1}^{N} K_i$, $H \equiv \sum_{i=1}^{N} H_i$, and $\Omega \equiv \sum_{i=1}^{N} \theta_{Ki}^{\alpha}\theta_{Hi}^{1-\alpha}\left(A_i/\overline{A}\right)$, with $\theta_{Ki} \equiv \frac{K_i}{K}$ and $\theta_{Hi} \equiv \frac{H_i}{H}$. Since (i) $\theta_{Ki}^{\alpha}\theta_{Hi}^{1-\alpha} \in [0,1]$ and (ii) $\sum_{i=1}^{N} \theta_{Ki}^{\alpha}\theta_{Hi}^{1-\alpha} \leq 1$, we can conclude $0 < \Omega \leq 1$.[34] Thus, to maximize Y given K and H, one should allocate all inputs to the most productive firm, when $\Omega = 1$ and $Y = \overline{A}K^{\alpha}H^{1-\alpha}$, which is expected as firms' goods are homogeneous. Therefore, $\Omega \in (0,1]$ represents the wedge between aggregate TFP, $\overline{A}\Omega$, and its optimal level, $\overline{A}$, being our measure of allocative efficiency.

Firms use inputs optimally, taking the same inputs' rental prices as given. As a consequence, from the FOCs of active firm i cost minimization problem,

$$\frac{w}{MPH_i} = \frac{r}{MPK_i} \leftrightarrow \frac{wH_i}{(1-\alpha)Y_i} = \frac{rK_i}{\alpha Y_i} \leftrightarrow \frac{K_i}{H_i} = \frac{w}{r}\frac{\alpha}{1-\alpha}, \tag{A.3}$$

where MPH_i and MPK_i are the marginal products of human and physical capital of active firm i, respectively. Thus, every active firm chooses the same physical-to-human capital ratio. Consequently, from (A.3), $K_i = \frac{K_j}{H_j}H_i$ for any firm i and active firm j, implying

$$\theta_{Kj} \equiv \frac{K_j}{K} = \frac{K_j}{\sum_{i=1}^{N} K_i} = \frac{K_j}{\sum_{i=1}^{N} \frac{K_j}{H_j} H_i} = \frac{H_j}{\sum_{i=1}^{N} H_i} = \frac{H_j}{L} \equiv \theta_{Hj}, \tag{A.4}$$

which allows us to define $\theta_j \equiv \theta_{Hj} = \theta_{Kj}$ for every firm j. Given that, one can use (A.3) to get

$$\alpha = \frac{K_i r}{K_i r + H_i w} = \frac{\theta_i K r}{\theta_i K r + \theta_i H w} = \frac{Kr}{Kr + Hw}, \tag{A.5}$$

[34] To see that $\sum_{i=1}^{N} \theta_{Ki}^{\alpha}\theta_{Hi}^{1-\alpha} \leq 1$, just use $\theta_{Ki}^{\alpha}\theta_{Hi}^{1-\alpha} = \exp\left(\alpha\ln(\theta_{Ki}) + (1-\alpha)\ln(\theta_{Hi})\right) \leq \exp\left[\ln(\alpha\theta_{Ki} + (1-\alpha)\theta_{Hi})\right] = \alpha\theta_{Ki} + (1-\alpha)\theta_{Hi}$, since ln is concave and exp is increasing.

meaning α equals the cost share of physical capital.

Using $\theta_{Ki} = \theta_{Hi} = \theta_i$, one can show that

$$s(A_i) = \frac{Y_i}{Y} = \frac{A_i K_i^{\alpha} H_i^{1-\alpha}}{\overline{A}\Omega K^{\alpha} H^{1-\alpha}} = \frac{A_i \theta_i}{\overline{A}\Omega} \rightarrow \theta_i = \overline{A}\Omega \frac{s(A_i)}{A_i} \tag{9}$$

$$\overline{A}\Omega = \frac{1}{\sum_{i=1}^{N} \frac{s(A_i)}{A_i}}, \tag{10}$$

where we use (2) and (8) to get (9), while (10) is obtained by summing (9) over all firms.

Finally, plugging Equations (6) and (7) into (10),

$$\Omega \approx \frac{\mathrm{E}_a \left[(\underline{A}/\overline{A})(1 - \underline{A}/A) \right]}{\mathrm{E}_a \left[(\underline{A}/A)(1 - \underline{A}/A) \right]}, \tag{11}$$

which has three interesting properties. First, it is easy to see that $\Omega \in (0, 1]$, as it should be given (8). Second, $\Omega \rightarrow 1$ as $\underline{A} \rightarrow \overline{A}$, which is an expected result since with no productivity dispersion, any allocation of resources is optimal. After all, from Equation (10),

$$1 = \sum_{i=1}^{N} s(A_i)(\overline{A}/\overline{A}) < \frac{1}{\Omega} = \sum_{i=1}^{N} s(A_i)(\overline{A}/A_i) < \sum_{i=1}^{N} s(A_i)(\overline{A}/\underline{A}) = \overline{A}/\underline{A}, \tag{A.6}$$

implying $1 \leq \lim_{\underline{A} \rightarrow \overline{A}} \Omega^{-1} \leq \lim_{\underline{A} \rightarrow \overline{A}} \overline{A}/\underline{A} = 1$. Note this result relies solely on the assumption of homogeneous goods and does not depend on the Cournot model. Third, the exit of less productive active firms leads to an improvement in Ω. To see that, we first assess the impact on allocative efficiency Ω resulting from the exit of only the least productive active firm (say, because η increases). Let $\underline{A} + \delta$, with $\delta > 0$, be the second-lowest productivity among active firms in the initial equilibrium. In this proof, we use the subscript 0 to denote the initial equilibrium and 1 for the final one, implying $\mathrm{E}_{a_0}(h(A)) \equiv \mathrm{E}(h(A) | A \geq \underline{A})$ and $\mathrm{E}_{a_1}(h(A)) \equiv \mathrm{E}(h(A) | A \geq \underline{A} + \delta)$. Moreover, from Equation (11),

$$\begin{aligned} \Omega_0 &\approx \frac{\mathrm{E}_{a_0} \left[(\underline{A}/\overline{A})(1 - \underline{A}/A) \right]}{\mathrm{E}_{a_0} \left[(\underline{A}/A)(1 - \underline{A}/A) \right]} = \frac{\mathrm{E} \left[(\underline{A}/\overline{A})(1 - \underline{A}/A) | A \geq \underline{A} \right]}{\mathrm{E} \left[(\underline{A}/A)(1 - \underline{A}/A) | A \geq \underline{A} \right]} \\ \Omega_0 &\approx \frac{\mathrm{E} \left[(\underline{A}/\overline{A})(1 - \underline{A}/A) | A \geq \underline{A} + \delta \right]}{\mathrm{E} \left[(\underline{A}/A)(1 - \underline{A}/A) | A \geq \underline{A} + \delta \right]} = \frac{\mathrm{E}_{a_1} \left[(\underline{A}/\overline{A})(1 - \underline{A}/A) \right]}{\mathrm{E}_{a_1} \left[(\underline{A}/A)(1 - \underline{A}/A) \right]} \end{aligned} \tag{A.7}$$

$$\Omega_1 \approx \frac{\mathrm{E}_{a_1} \left[\left((\underline{A} + \delta)/\overline{A} \right) \left(1 - (\underline{A} + \delta)/A \right) \right]}{\mathrm{E}_{a_1} \left[\left((\underline{A} + \delta)/A \right) \left(1 - (\underline{A} + \delta)/A \right) \right]}, \tag{A.8}$$

where we use $1 - \underline{A}/\underline{A} = 0$ in the second line. Therefore, $\Delta\Omega \equiv \Omega_1 - \Omega_0$ is

$$\Delta\Omega \approx \frac{\mathrm{E}_{a_1} \left[\left((\underline{A} + \delta)/\overline{A} \right) \left(1 - (\underline{A} + \delta)/A \right) \right]}{\mathrm{E}_{a_1} \left[\left((\underline{A} + \delta)/A \right) \left(1 - (\underline{A} + \delta)/A \right) \right]} - \frac{\mathrm{E}_{a_1} \left[(\underline{A}/\overline{A})(1 - \underline{A}/A) \right]}{\mathrm{E}_{a_1} \left[(\underline{A}/A)(1 - \underline{A}/A) \right]}$$

$$
\begin{aligned}
\Delta\Omega \approx & \frac{\mathrm{E}_{a_1}\left[\left(\underline{A}+\delta\right)\left(\frac{A-(\underline{A}+\delta)}{\overline{A}A}\right)\right]}{\mathrm{E}_{a_1}\left[\left(\underline{A}+\delta\right)\left(\frac{A-(\underline{A}+\delta)}{A^2}\right)\right]} - \frac{\mathrm{E}_{a_1}\left[\underline{A}\left(\frac{A-\underline{A}}{\overline{A}A}\right)\right]}{\mathrm{E}_{a_1}\left[\underline{A}\left(\frac{A-\underline{A}}{A^2}\right)\right]} = \frac{\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{\overline{A}A}-\frac{\delta}{\overline{A}A}\right)}{\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}-\frac{\delta}{A^2}\right)} - \frac{\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{\overline{A}A}\right)}{\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)} \\
\Delta\Omega \approx & \frac{\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{\overline{A}A}-\frac{\delta}{\overline{A}A}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)-\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{\overline{A}A}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}-\frac{\delta}{A^2}\right)}{\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}-\frac{\delta}{A^2}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)} \\
\Delta\Omega \approx & \frac{-\mathrm{E}_{a_1}\left(\frac{\delta}{\overline{A}A}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)+\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{\overline{A}A}\right)\mathrm{E}_{a_1}\left(\frac{\delta}{A^2}\right)}{\mathrm{E}_{a_1}\left(\frac{A-(\underline{A}+\delta)}{A^2}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)} \\
\Delta\Omega \approx & \frac{\mathrm{E}_{a_1}\left(1-\frac{\underline{A}}{A}\right)\mathrm{E}_{a_1}\left(\frac{1}{A^2}\right)-\mathrm{E}_{a_1}\left(\frac{1}{A}\right)\mathrm{E}_{a_1}\left(\frac{1}{A}-\frac{\underline{A}}{A^2}\right)}{(\overline{A}/\delta)\mathrm{E}_{a_1}\left(\frac{A-(\underline{A}+\delta)}{A^2}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)} \\
\Delta\Omega \approx & \frac{-\mathrm{E}_{a_1}\left(\frac{1}{A}\right)^2+\underline{A}\mathrm{E}_{a_1}\left(\frac{1}{A}\right)\mathrm{E}_{a_1}\left(\frac{1}{A^2}\right)+\mathrm{E}_{a_1}\left(\frac{1}{A^2}\right)-\underline{A}\mathrm{E}_{a_1}\left(\frac{1}{A}\right)\mathrm{E}_{a_1}\left(\frac{1}{A^2}\right)}{(\overline{A}/\delta)\mathrm{E}_{a_1}\left(\frac{A-(\underline{A}+\delta)}{A^2}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)} \\
\Delta\Omega \approx & \frac{\mathrm{Var}_{a_1}\left(\frac{1}{A}\right)}{(\overline{A}/\delta)\mathrm{E}_{a_1}\left(\frac{A-(\underline{A}+\delta)}{A^2}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)},
\end{aligned}
\tag{A.9}
$$

where $\mathrm{Var}_{a_1}(h(A)) \equiv \mathrm{Var}(h(A)|A \geq \underline{A}+\delta) = \mathrm{E}_{a_1}(h(A)^2) - [\mathrm{E}_{a_1}(h(A))]^2$ is the variance of $h(A)$ over active firms under the empirical distribution in the final equilibrium, for any function h. Note our analysis implicitly assumes the presence of productivity dispersion in the initial equilibrium. Within the initial set of active firms, the second-lowest productivity level is $\underline{A}+\delta$, which is strictly greater than the lowest level ($\underline{A}$) as $\delta > 0$. As a result, productivity dispersion should also be present in the final equilibrium. After all, if this were not the case, $\mathrm{E}_{a_1}(\underline{A}/A) = 1$, implying from (7) that either $N_{a_1} = +\infty$ or $\eta_1 = +\infty$. While $\eta_1 = +\infty$ is ruled out by assumption, the scenario of $N_{a_1} = +\infty$ cannot hold since there is productivity dispersion in the initial equilibrium. All in all, we can conclude $\mathrm{Var}_{a_1}(1/A) > 0$ and consequently $\Delta\Omega > 0$. Applying this result iteratively, we would conclude the exit of less productive active firms improves the allocative efficiency Ω.

A.5 Average markup

Using this model, we can also compute the cost-weighted average of firm-level markups $\mu \equiv \sum_{i=1}^{N}\left(\frac{H_i w + K_i r}{Hw+Kr}\right)\mu_i = \sum_{i=1}^{N}\theta_i\mu_i$ through

$$
\mu = \sum_{i=1}^{N}\left[\overline{A}\Omega\frac{s(A_i)}{A_i}\right]\left(\frac{p}{MC_i}\right) \approx \sum_{i=1}^{N}\left[\overline{A}\Omega\frac{s(A_i)}{A_i}\right]\left[\frac{\left(\frac{r}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\frac{1}{\underline{A}}}{\left(\frac{r}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\frac{1}{A_i}}\right] = \frac{\overline{A}\Omega}{\underline{A}}, \tag{12}
$$

where we use (9), (4), that the least productive active firm has markup approximately equal to one, and the Cobb-Douglas marginal cost function.

B Derivation of the continuous model

Given the similarity to the discrete model, we refrain from presenting the derivation in detail. Instead, our focus is on emphasizing the main differences from the discrete case. To derive the model, the SOC of firms' profit maximization must hold. In the discrete case, one can easily see this condition is satisfied for $\eta > 1$. However, in the continuous model, such evaluation is less straightforward, as it requires considering the equilibrium values of s_i. As a consequence, in deriving the model, we simply assume this condition is met. We subsequently validate this claim in Appendix B.6 using the model solution for s_i, demonstrating the SOC holds if $\eta > 0$ or if $q \in (0, 1]$ is low.

B.1 Environment and technology

The absolute value of the price elasticity of demand is η, with $0 < \eta < +\infty$. Since now there is a continuum of firms $i \in [0, N]$, the aggregate output is given by an integral instead of a sum:

$$Y \equiv \int_0^N Y_i di, \tag{1c}$$

where Y_i is still given by (2).

B.2 Market competition and optimal decision

Firms' problem continues to be (3). Consequently, the FOC is still given by (4), but now $\eta_i = \frac{\eta/q}{s_i}$ since $\partial Y/\partial Y_i = q \in (0, 1]$ instead of $\partial Y/\partial Y_i = 1$. This means the markup of firm i becomes $\mu_i = \left(1 - \frac{s_i}{\eta/q}\right)^{-1} = \frac{\eta/q}{\eta/q - s_i}$. Similarly, the SOC is also different, given by

$$\begin{aligned}
0 >& \frac{\partial p}{\partial Y}\frac{\partial Y}{\partial Y_i}\left(1 - \frac{s_i}{\eta/q}\right) - \frac{p}{\eta/q}\left(\frac{Y - Y_i\frac{\partial Y}{\partial Y_i}}{Y^2}\right) \\
0 >& -\frac{p}{(\eta/q)Y}\left(1 - \frac{s_i}{\eta/q}\right) - \frac{p}{(\eta/q)Y}(1 - s_i q) = -\frac{p}{(\eta^2/q)Y}\left[2\eta - s_i q(1+\eta)\right] \\
s_i q <& \frac{2\eta}{1+\eta}.
\end{aligned} \tag{A.1c}$$

For now, simply assume the SOC holds, implying (4) represents the optimal decision for all active firms, that is, for every firm i such as $s_i \geq 0$ or, equivalently, $\mu_i \geq 1$.

B.3 Equilibrium allocation

Equation (5) is no longer valid. However, from (4) with $\eta_i = \frac{\eta/q}{s_i}$, one can easily show that

$$s_i - s_j = \left(1 - \frac{A_j}{A_i}\right)(\eta/q - s_j), \tag{5c}$$

which defines s_i as a strictly increasing function of A_i, since $\eta/q - s_j > 0$ as $\mu_j = \frac{\eta/q}{\eta/q - s_j} > 1$ for any active firm j. As before, if some firms may be inactive, we seek the unique equilibrium in which a firm i is active if and only if $A_i \geq \underline{A}$ for some productivity cutoff $\underline{A}$.

Given that, one can use (5c) to obtain, analogously to (A.2),

$$s(A_i) = \frac{1}{N_a \mathrm{E}_a(A_i/A)} + (\eta/q)\left(1 - \frac{1}{\mathrm{E}_a(A_i/A)}\right), \tag{A.2c}$$

where $\mathrm{E}_a\left(h(A)\right) \equiv \mathrm{E}\left(h(A)|A \geq \underline{A}\right) = \int_{\underline{A}}^{\overline{A}} h(A)\frac{g(A)}{1-G(\underline{A})}dA$ is the expected value of a function h over active firms under the empirical distribution, $g(A) > 0$ is the empirical density of A, and $G(A) = \int_{\underline{\underline{A}}}^{A} g(a)da$ is the empirical cumulative distribution function. As before, $N_a \equiv N\left(1 - G(\underline{A})\right)$ is the number of active firms. Using (A.2c), let us show not all firms can be active simultaneously if N is sufficiently large. Assume, by contradiction, all firms are active, with $N_a = N \to +\infty$. From (A.2c), it is easy to see $s(\underline{A}) < 0$ under such circumstance unless $E_a(\underline{A}/A) \to 1$ and thus $\underline{A} \to \overline{A}$, when you would get the perfect competition case. However, this could not hold as all firms are active and $A_i \in [\underline{\underline{A}}, \overline{A}]$, with $\underline{\underline{A}} < \overline{A}$.

Therefore, assuming N is sufficiently large, some low-productivity firms would be inactive. Consequently, $s(\underline{A}) = 0$. To see this last result, it is sufficient to show such $\underline{A}$ exists. After all, on the one hand, if it exists, $s(\underline{A}) > 0$ cannot be an equilibrium given our assumption of free entry. On the other hand, $s(\underline{A}) < 0$ is never an equilibrium due to free exit. A productivity level $\underline{A}$ such as $s(\underline{A}) = 0$ exists because (i) $s(\underline{A}) < 0$ as $\underline{A} \to \underline{\underline{A}}$, (ii) $s(\underline{A}) > 0$ as $\underline{A} \to \overline{A}$, and (iii) $s(\underline{A})$ is continuous in $\underline{A}$. The first result holds because s_i is strictly increasing in A_i and, by assumption, the mass N of firms is large, meaning not all firms can be active simultaneously. The last two results can be obtained from (A.2c). From this expression, $s(\underline{A}) \to 1/N_a = 1/[N(1 - G(\underline{A}))] \to +\infty$ as $\underline{A} \to \overline{A}$, yielding the second result. The third result holds as both $N_a \equiv N(1 - G(\underline{A}))$ and $\mathrm{E}_a(1/A)$ are continuous functions of $\underline{A}$, since $\mathrm{E}_a(1/A)$ is differentiable in $\underline{A}$ (Proposition D.1).

Finally, plugging $s(\underline{A}) = 0$ respectively into (5c) and (A.2c), note

$$s(A_i) = (\eta/q)\left(1 - \underline{A}/A_i\right) \tag{6c}$$

$$\eta/q = \frac{1}{N_a\left[1 - \mathrm{E}_a\left(\underline{A}/A\right)\right]}, \tag{7c}$$

which are the continuous versions of (6) and (7), respectively.

B.4 Aggregate productivity and misallocation

Equation (8) remains valid with $K \equiv \int_0^N K_i di$, $H \equiv \int_0^N H_i di$, and $\Omega \equiv \int_0^N \theta_{Ki}^{\alpha}\theta_{Hi}^{1-\alpha}\left(A_i/\overline{A}\right) di$. Equations (A.3), (A.4), (A.5), and (9) also remain valid. Naturally, (10) should be replaced by

$$\overline{A}\Omega = \frac{1}{\int_0^N \frac{s(A_i)}{A_i} di}. \tag{10c}$$

Finally, allocative efficiency Ω continues to be given by (11), but now holding *exactly*. Thus, it shows similar properties. First, as can be easily seen, $\Omega \in (0, 1]$. Second, $\Omega \to 1$ as $\underline{A} \to \overline{A}$, since, from (10c),

$$1 = \int_0^N s(A_i)(\overline{A}/\overline{A})di < \frac{1}{\Omega} = \int_0^N s(A_i)(\overline{A}/A_i)di < \int_0^N s(A_i)(\overline{A}/\underline{A})di = \overline{A}/\underline{A}. \tag{A.6c}$$

Third, Ω is strictly increasing in $\underline{A}$, since, from Equation (11) holding exactly,

$$\begin{aligned}\frac{\partial\Omega}{\partial\underline{A}} =& \frac{\left[-\mathrm{E}_a(1/A) - \underline{A}\frac{\partial \mathrm{E}_a(1/A)}{\partial \underline{A}}\right]\left[\mathrm{E}_a(1/A) - \underline{A}\mathrm{E}_a(1/A^2)\right]}{\overline{A}\left[\mathrm{E}_a(1/A) - \underline{A}\mathrm{E}_a(1/A^2)\right]^2} \\ &- \frac{\left[1 - \underline{A}\mathrm{E}_a(1/A)\right]\left[\frac{\partial \mathrm{E}_a(1/A)}{\partial \underline{A}} - E(1/A^2) - \underline{A}\frac{\partial \mathrm{E}_a(1/A^2)}{\partial \underline{A}}\right]}{\overline{A}\left[\mathrm{E}_a(1/A) - \underline{A}\mathrm{E}_a(1/A^2)\right]^2}\end{aligned}$$

$$\begin{aligned}\frac{\partial\Omega}{\partial\underline{A}} =& \frac{\frac{\partial \mathrm{E}_a(1/A)}{\partial \underline{A}}\left[\underline{A}^2\mathrm{E}_a(1/A^2) - \underline{A}\mathrm{E}_a(1/A) - 1 + \underline{A}\mathrm{E}_a(1/A)\right] + \underline{A}\frac{\partial \mathrm{E}_a(1/A^2)}{\partial \underline{A}}\left[1 - \underline{A}\mathrm{E}_a(1/A)\right]}{\overline{A}\left[\mathrm{E}_a(1/A) - \underline{A}\mathrm{E}_a(1/A^2)\right]^2} \\ &- \frac{\mathrm{E}_a(1/A)\left[\mathrm{E}_a(1/A) - \underline{A}\mathrm{E}_a(1/A^2)\right] - \left[1 - \underline{A}\mathrm{E}_a(1/A)\right]E(1/A^2)}{\overline{A}\left[\mathrm{E}_a(1/A) - \underline{A}\mathrm{E}_a(1/A^2)\right]^2}\end{aligned}$$

$$\begin{aligned}\frac{\partial\Omega}{\partial\underline{A}} =& \frac{\frac{\partial \mathrm{E}_a(1/A)}{\partial \underline{A}}\left\{\mathrm{E}_a\left[(\underline{A}/A)^2\right] - 1\right\} - \underline{A}\frac{\partial \mathrm{E}_a(1/A^2)}{\partial \underline{A}}\left[\mathrm{E}_a(\underline{A}/A) - 1\right]}{\overline{A}\left[\mathrm{E}_a(1/A) - \underline{A}\mathrm{E}_a(1/A^2)\right]^2} \\ &+ \frac{\mathrm{E}_a(1/A^2) - \left[\mathrm{E}_a(1/A)\right]^2}{\overline{A}\left[\mathrm{E}_a(1/A) - \underline{A}\mathrm{E}_a(1/A^2)\right]^2}\end{aligned}$$

$$\begin{aligned}\frac{\partial\Omega}{\partial\underline{A}} =& \frac{\tilde{g}(\underline{A})\left[\mathrm{E}_a(1/A) - (1/\underline{A})\right]\left\{\mathrm{E}_a\left[(\underline{A}/A)^2\right] - 1\right\} - \underline{A}\tilde{g}(\underline{A})\left[\mathrm{E}_a(1/A^2) - (1/\underline{A}^2)\right]\left[\mathrm{E}_a(\underline{A}/A) - 1\right]}{\overline{A}\left[\mathrm{E}_a(1/A) - \underline{A}\mathrm{E}_a(1/A^2)\right]^2} \\ &+ \frac{\mathrm{Var}_a(1/A)}{\overline{A}\left[\mathrm{E}_a(1/A) - \underline{A}\mathrm{E}_a(1/A^2)\right]^2}\end{aligned}$$

$$\frac{\partial\Omega}{\partial\underline{A}} = \left[\frac{\tilde{g}(\underline{A})}{\underline{A}}\right]\frac{\left[\mathrm{E}_a(\underline{A}/A) - 1\right]\left\{\mathrm{E}_a\left[(\underline{A}/A)^2\right] - 1\right\} - \left\{\mathrm{E}_a\left[(\underline{A}/A)^2\right] - 1\right\}\left[\mathrm{E}_a(\underline{A}/A) - 1\right]}{\overline{A}\left[\mathrm{E}_a(1/A) - \underline{A}\mathrm{E}_a(1/A^2)\right]^2}$$

$$+\frac{\mathrm{Var}_a(\underline{A}/A)}{\overline{A}\left\{\mathrm{E}_a(\underline{A}/A)-\mathrm{E}_a\left[(\underline{A}/A)^2\right]\right\}^2}$$

$$\frac{\partial\Omega}{\partial\underline{A}}=\frac{\mathrm{Var}_a(\underline{A}/A)}{\overline{A}\left\{\mathrm{E}_a(\underline{A}/A)-\mathrm{E}_a\left[(\underline{A}/A)^2\right]\right\}^2},$$

where $\mathrm{Var}_a(h(A))\equiv\mathrm{E}_a\left(h(A)^2\right)-\left[\mathrm{E}_a(h(A))\right]^2$ and we use Proposition D.1 to get $\frac{\partial\mathrm{E}_a(1/A)}{\partial\underline{A}}$ and $\frac{\partial\mathrm{E}_a(1/A^2)}{\partial\underline{A}}$ in the fourth line. Hence, $\frac{\partial\Omega}{\partial\underline{A}}>0$, since $\mathrm{Var}_a\left(1/A\right)>0$ given (7c) with $0<\eta<+\infty$ and $q\in(0,1]$.

B.5 Average markup

Equation (12) is now *exactly* valid.

B.6 Assessing the SOC for profit maximization

As our final task, we assess if the SOC for profit maximization (A.1c) holds. Note $\eta>1$ is not sufficient because, with a continuum of firms, s_i is a density function, and consequently, it may be strictly greater than one. Since s_i is an endogenous variable, let us evaluate this condition in the model, using $s_iq=\eta\left(1-\underline{A}/A_i\right)$ from Equation (6c).

Formally, we need to demonstrate that $\frac{2\eta}{1+\eta}>s_iq=\eta\left(1-\underline{A}/A_i\right)$ for every active firm i. We address this issue under two cases: (i) $\eta\in(0,1]$ and (ii) $\eta>1$. On the one hand, since $\eta\in(0,1]\rightarrow\frac{2\eta}{1+\eta}\geq\eta$ and $\eta>s_iq$ as $\mu_i=\frac{\eta}{\eta-s_iq}>1$ for any active firm i, the SOC holds for any $\eta,q\in(0,1]$.[35] On the other hand, the SOC holds for $\eta\approx1$, $\eta>1$, since it holds for $\eta=1$ and we are just dealing with continuous functions of η, but it is not fulfilled for high enough η. To get this last result, use (7c) to see that $\underline{A}\rightarrow\overline{A}$ as $\eta\rightarrow+\infty$, and thus $\eta\mathrm{E}_a(1-\underline{A}/A)=\frac{q}{N_a}=\frac{q}{N(1-G(\underline{A}))}\rightarrow+\infty$. As a consequence, given that $(1-\underline{A}/\overline{A})\geq\mathrm{E}_a(1-\underline{A}/A)$, $\eta(1-\underline{A}/\overline{A})\rightarrow+\infty$ as $\eta\rightarrow+\infty$. This implies the SOC will not hold for the most productive firms under high η since the LHS of (A.1c) is always lower than 2. Moreover, with $\eta>1$, the SOC holds for low $q>0$, since $s_iq=\eta\left(1-\underline{A}/A_i\right)\rightarrow0$ as $q\rightarrow0^+$, because, from (7c), $\underline{A}\rightarrow\overline{A}$ under such condition.

In short, the SOC holds for low $\eta>0$ or low $q\in(0,1]$. By choosing both η and q, we can simultaneously fulfill such condition and choose the adjusted elasticity of demand η/q (e.g., set $\eta=1$ and choose q to get the desired η/q). However, in empirical applications, calibrating η/q is not always necessary, as several key variables, such as allocative efficiency Ω, can be computed without it. In this context, one may only assume $\eta>0$ and $q\in(0,1]$ are such that $\frac{2\eta}{1+\eta}>s_iq$ for every active firm i. We adopt this approach in the main text.

[35] Through analogous reasoning, it can be shown that $\eta\in(0,1]$ is also suitable for the discrete model. Therefore, while $\eta>1$ is a sufficient condition there, it is not necessary.

C Microfoundation of the price elasticity of demand

To microfoundate the price elasticity of demand η, rather than relying on a single-sector (or single-good) economy, we examine a multiple-sector economy that can, however, be represented within a single-sector environment. Under this new setup, our model would apply to the representative sector. However, in this case, the price elasticity of demand is not an ad hoc parameter; instead, it is equal to the elasticity of substitution across sectors.

Formally, suppose a representative perfectly competitive firm produces a homogeneous final good Y using inputs Y_s from a continuum of sectors through the CES production function

$$Y = \left(\int_0^1 Y_s^{\frac{\sigma-1}{\sigma}} ds\right)^{\frac{\sigma}{\sigma-1}}, \tag{C.1}$$

where σ is the elasticity of substitution across sectors $s \in [0, 1]$. Therefore, if the final good is the numeraire, this representative firm solves the profit maximization problem

$$\begin{aligned} \max_{\{Y_s\}_{s\in[0,1]}} \quad & \left(\int_0^1 Y_s^{\frac{\sigma-1}{\sigma}} ds\right)^{\frac{\sigma}{\sigma-1}} - \int_0^1 Y_s p_s ds \\ \text{s.t.} \quad & p_s > 0, \quad \forall s \in [0,1] \end{aligned}, \tag{C.2}$$

with p_s representing the price of goods within sector s, which are homogeneous across all firms in that sector. The FOCs of this problem are

$$p_s = Y^{1/\sigma} Y_s^{-1/\sigma}, \quad \forall s \in [0, 1]. \tag{C.3}$$

In each sector, intermediate firms make their decisions taking the price and quantity of the final good as given.[36] As a result, being η_s the absolute value of the price elasticity of demand for the good of sector s, using (C.3), we can see that

$$\eta_s \equiv -\left(\frac{\partial p_s}{\partial Y_s}\frac{Y_s}{p_s}\right)^{-1} = \sigma\left(Y^{1/\sigma}Y_s^{-1/\sigma-1}\frac{Y_s}{Y^{1/\sigma}Y_s^{-1/\sigma}}\right)^{-1} = \sigma. \tag{C.4}$$

Hence, the price elasticity of demand is the same for all sectors, given by the elasticity of substitution σ. If we further assume firms' technology, the number of firms, and the distribution of productivity are identical across sectors, each sector would face the same problem, and consequently $Y_s = Y_{\tilde{s}}$ for any sectors $s, \tilde{s} \in [0, 1]$. Plugging it into (C.1),

$$Y = \left(\int_0^1 Y_{\tilde{s}}^{\frac{\sigma-1}{\sigma}} ds\right)^{\frac{\sigma}{\sigma-1}} = Y_{\tilde{s}}, \tag{C.5}$$

[36] This assumption is also pivotal if one aims to specify and solve a full macroeconomic model because it overcomes the technical problems associated with embedding oligopoly models into general equilibrium frameworks (see Neary 2010 for a discussion). This represents another advantage of this alternative interpretation of the model.

implying we can treat the economy as if there were a single sector, which we analyze through the lens of a static Cournot model.

D Conditions for the calibration algorithm to work properly

Owing to the simplicity of our model, we can establish necessary and sufficient conditions for the calibration algorithm of Section 3.3 to work properly, achieving an exact match of both target moments. We need to show that a unique solution exists for its first-step problem or, equivalently, that (14) implicitly defines $\tilde{A} \equiv \overline{A}/\underline{A}$ as a well-defined function of μ. Initially, we derive some general results by examining an arbitrary continuous truncated distribution. The only requirement is that its density can be expressed as an upper truncation of another distribution. These general findings provide a framework for establishing conditions applicable to any such distribution. In the latter part, we utilize this framework to delineate the specific conditions pertaining to the Pareto case. Additionally, in the final proposition of this section, we assess what happens with the estimated Ω under the Pareto distribution as $k \to -\infty$.

In the following, consider μ as given in (14), that is,

$$\mu = \frac{1 - \mathrm{E}_a\left(\underline{A}/A\right)}{\mathrm{E}_a\left(\underline{A}/A\right) - \mathrm{E}_a\left[(\underline{A}/A)^2\right]} = \frac{\mathrm{E}_a\left(1 - \underline{A}/A\right)}{\mathrm{E}_a\left(1 - \underline{A}/A\right) - \mathrm{E}_a\left[(1 - \underline{A}/A)^2\right]}, \tag{14}$$

where $\mathrm{E}_a\left(h(A)\right) \equiv \mathrm{E}\left(h(A) | \underline{A} \le A \le \overline{A}\right)$ for any function h.

D.1 Arbitrary distribution

Assume $A \in [\underline{\underline{A}}, \overline{A}]$ is a continuous variable, $0 < \underline{\underline{A}} < \overline{A} < +\infty$, whose density and cumulative distribution function are g and G, respectively. Let $\tilde{g}(A) \equiv \frac{g(A)}{1-G(\underline{A})} > 0$, $\underline{A} \in (\underline{\underline{A}}, \overline{A})$, the density function of $A \in [\underline{A}, \overline{A}]$, with cumulative distribution $\tilde{G}$. Let $\hat{g}$ be another density of A, but defined over the support $A \in [\underline{A}, A_h]$, $A_h > \overline{A}$, possibly with $A_h \to +\infty$. This density does not depend on $\overline{A}$ and has cumulative distribution function $\hat{G}$. Moreover, it satisfies $\tilde{g}(A) = \frac{\hat{g}(A)}{\hat{G}(\overline{A})}$, meaning $\tilde{g}$ is an upper truncation of $\hat{g}$.

Proposition D.1 *Let h be a function of A for which $\mathrm{E}_a\left(h(A)\right)$ is well defined. In this case,*

1. *If* $\frac{\partial h(A)}{\partial \underline{A}} = 0$, $\frac{\partial \mathrm{E}_a(h(A))}{\partial \underline{A}} = \tilde{g}(\underline{A})\left[\mathrm{E}_a\left(h(A)\right) - h(\underline{A})\right]$.
2. *If* $\frac{\partial h(A)}{\partial \overline{A}} = 0$, $\frac{\partial \mathrm{E}_a(h(A))}{\partial \overline{A}} = \tilde{g}(\overline{A})\left[h(\overline{A}) - \mathrm{E}_a\left(h(A)\right)\right]$.

Proof. Let h be a function of A for which $\mathrm{E}_a\left(h(A)\right)$ is well defined. If $\frac{\partial h(A)}{\partial \underline{A}} = 0$,

$$\begin{aligned}
\frac{\partial \mathrm{E}_a\left(h(A)\right)}{\partial \underline{A}} &= \frac{\partial \left[\frac{\int_{\underline{A}}^{\overline{A}} h(A)g(A)dA}{1-G(\underline{A})}\right]}{\partial \underline{A}} = \frac{-h(\underline{A})g(\underline{A})\left(1 - G(\underline{A})\right) + \left[\int_{\underline{A}}^{\overline{A}} h(A)g(A)dA\right] g(\underline{A})}{\left[1 - G(\underline{A})\right]^2} \\
\frac{\partial \mathrm{E}_a\left(h(A)\right)}{\partial \underline{A}} &= \tilde{g}(\underline{A})\left[\mathrm{E}_a\left(h(A)\right) - h(\underline{A})\right],
\end{aligned}$$

where we use $\frac{\partial h(A)}{\partial \underline{A}} = 0$ to get $\frac{\partial\left[\int_{\underline{A}}^{\overline{A}} h(A)g(A)dA\right]}{\partial \underline{A}} = -h(\underline{A})g(\underline{A})$. If $\frac{\partial h(A)}{\partial \overline{A}} = 0$,

$$\frac{\partial \mathrm{E}_a\left(h(A)\right)}{\partial \overline{A}} = \frac{\partial\left[\frac{\int_{\underline{A}}^{\overline{A}} h(A)\hat{g}(A)dA}{\hat{G}(\overline{A})}\right]}{\partial \overline{A}} = \frac{h(\overline{A})\hat{g}(\overline{A})\hat{G}(\overline{A}) - \left[\int_{\underline{A}}^{\overline{A}} h(A)\hat{g}(A)dA\right]\hat{g}(\overline{A})}{\hat{G}(\overline{A})^2}$$
$$\frac{\partial \mathrm{E}_a\left(h(A)\right)}{\partial \overline{A}} = \tilde{g}(\overline{A})\left[h(\overline{A}) - \mathrm{E}_a\left(h(A)\right)\right],$$

where we use $\frac{\partial h(A)}{\partial \overline{A}} = 0$ to get $\frac{\partial\left[\int_{\underline{A}}^{\overline{A}} h(A)\hat{g}(A)dA\right]}{\partial \overline{A}} = h(\overline{A})\hat{g}(\overline{A})$. ■

Proposition D.2 $\frac{\partial \mu}{\partial \overline{A}} > 0$.

Proof. From Equation (14),

$$\frac{\partial \mu}{\partial \overline{A}} = \frac{\frac{\partial \mathrm{E}_a(1-\underline{A}/A)}{\partial \overline{A}}\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2} - \frac{\mathrm{E}_a\left(1-\underline{A}/A\right)\left\{\frac{\partial \mathrm{E}_a(1-\underline{A}/A)}{\partial \overline{A}} - \frac{\partial \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{\partial \overline{A}}\right\}}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2}$$

$$\frac{\partial \mu}{\partial \overline{A}} = \frac{\mathrm{E}_a\left(1-\underline{A}/A\right)\frac{\partial \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{\partial \overline{A}} - \frac{\partial \mathrm{E}_a(1-\underline{A}/A)}{\partial \overline{A}}\mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2}$$

$$\frac{\partial \mu}{\partial \overline{A}} = \frac{\mathrm{E}_a\left(1-\underline{A}/A\right)\tilde{g}(\overline{A})\left\{(1-\underline{A}/\overline{A})^2 - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2} - \frac{\tilde{g}(\overline{A})\left[(1-\underline{A}/\overline{A}) - \mathrm{E}_a\left(1-\underline{A}/A\right)\right]\mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2}$$

$$\frac{\partial \mu}{\partial \overline{A}} = \frac{\tilde{g}(\overline{A})(1-\underline{A}/\overline{A})^2\mathrm{E}_a\left(1-\underline{A}/A\right) - \tilde{g}(\overline{A})(1-\underline{A}/\overline{A})\mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2}$$

$$\frac{\partial \mu}{\partial \overline{A}} = \frac{\tilde{g}(\overline{A})(1-\underline{A}/\overline{A})\mathrm{E}_a\left[(1-\underline{A}/A)\left(\underline{A}/A - \underline{A}/\overline{A}\right)\right]}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2},$$

where we use Proposition D.1 to get $\frac{\partial \mathrm{E}_a(1-\underline{A}/A)}{\partial \overline{A}}$ and $\frac{\partial \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{\partial \overline{A}}$ in the third line. Therefore, $\frac{\partial \mu}{\partial \overline{A}} > 0$ as $(1-\underline{A}/A)\left(\underline{A}/A - \underline{A}/\overline{A}\right) > 0$ for $A \in (\underline{A}, \overline{A})$. ■

Proposition D.3 $\lim_{\overline{A} \to \underline{A}^+} \mu = 1$.

Proof. From Equation (14),

$$1 = \frac{\mathrm{E}_a\left(1 - \underline{A}/A\right)}{\mathrm{E}_a\left[(1 - \underline{A}/A)(\underline{A}/\underline{A})\right]} \leq \mu = \frac{\mathrm{E}_a\left(1 - \underline{A}/A\right)}{\mathrm{E}_a\left[(1 - \underline{A}/A)(\underline{A}/A)\right]} \leq \frac{\mathrm{E}_a\left(1 - \underline{A}/A\right)}{\mathrm{E}_a\left[(1 - \underline{A}/A)(\underline{A}/\overline{A})\right]} = \overline{A}/\underline{A}.$$

Therefore, $1 \leq \mu \leq \overline{A}/\underline{A}$, implying $\lim_{\overline{A}\to\underline{A}^+} \mu = 1$. ■

Proposition D.4 *$\overline{A}$, $\overline{A} > \underline{A}$, is a continuous, strictly increasing, and well-defined function of μ if and only if $\mu \in \left(1, \lim_{\overline{A}\to+\infty} \mu\right)$.*

Proof. First, μ is continuous and strictly increasing in $\overline{A}$, $\overline{A} > \underline{A}$, as it is differentiable with $\frac{\partial\mu}{\partial\overline{A}} > 0$ (Proposition D.2), implying it is an one-to-one function. Second, since $\lim_{\overline{A}\to\underline{A}^+} \mu = 1$ (Proposition D.3), the image of μ over $\overline{A} \in (\underline{A}, +\infty)$ is $\left(1, \lim_{\overline{A}\to+\infty} \mu\right)$ as μ is continuous and strictly increasing in $\overline{A}$. As a consequence of these two results, $\overline{A}$, $\overline{A} > \underline{A}$, is a continuous, strictly increasing, and well-defined function of μ if and only if $\mu \in \left(1, \lim_{\overline{A}\to+\infty} \mu\right)$. ■

D.2 Pareto distribution of firm productivity

If $A \in \left[\underline{A}, \overline{A}\right] \in (0, +\infty)$ is truncated Pareto distributed with shape parameter $k \neq 0$, its density is $g(A) = k\left(\frac{\underline{A}^k\overline{A}^k}{\overline{A}^k - \underline{A}^k}\right) A^{-k-1}$ and its cumulative distribution function is $G(A) = \frac{\overline{A}^k - \underline{A}^k\overline{A}^k A^{-k}}{\overline{A}^k - \underline{A}^k}$. Note this density is well defined for any $k \neq 0$ as $g(A) > 0\ \forall A \in \left[\underline{A}, \overline{A}\right]$ and $G(\overline{A}) = 1$. Moreover, it is easy to see $A \in \left[\underline{A}, \overline{A}\right]$ has also a truncated Pareto distribution with parameter $k \neq 0$, since its density is $\tilde{g}(A) \equiv \frac{g(A)}{1 - G(\underline{A})} = k\left(\frac{\underline{A}^k\overline{A}^k}{\overline{A}^k - \underline{A}^k}\right) A^{-k-1}$. Consider in the following $\tilde{A} \equiv \overline{A}/\underline{A} > 1$.

Proposition D.5 *For $k \neq 0$ and $j \in \mathbb{N}\backslash\{0\}$,* $\mathrm{E}_a\left((\underline{A}/A)^j\right) = \begin{cases} \left(\frac{k}{k+j}\right)\left(\frac{\tilde{A}^{k+j} - 1}{\tilde{A}^{k+j} - \tilde{A}^j}\right) & \textit{, if } k + j \neq 0 \\ \left(\frac{k\tilde{A}^k}{\tilde{A}^k - 1}\right) \ln\tilde{A} & \textit{, if } k + j = 0 \end{cases}.$

Proof. Let $k \neq 0$ and $j \in \mathbb{N} \setminus \{0\}$. From the truncated Pareto density, if $k + j \neq 0$,

$$\mathrm{E}_a\left((\underline{A}/A)^j\right) = \int_{\underline{A}}^{\overline{A}} k\left(\frac{\underline{A}^{k+j}\overline{A}^k}{\overline{A}^k - \underline{A}^k}\right) A^{-k-1-j} dA = k\left(\frac{\underline{A}^{k+j}\overline{A}^k}{\overline{A}^k - \underline{A}^k}\right)\left(\frac{\overline{A}^{-k-j} - \underline{A}^{-k-j}}{-k-j}\right)$$

$$\mathrm{E}_a\left((\underline{A}/A)^j\right) = k\left(\frac{\tilde{A}^k}{\tilde{A}^k - 1}\right)\left(\frac{\tilde{A}^{-k-j} - 1}{-k-j}\right) = \left(\frac{k}{k+j}\right)\left(\frac{\tilde{A}^{k+j} - 1}{\tilde{A}^{k+j} - \tilde{A}^j}\right),$$

while, if $k + j = 0$, $\mathrm{E}_a\left((\underline{A}/A)^j\right) = \int_{\underline{A}}^{\overline{A}} k\left(\frac{\overline{A}^k}{\overline{A}^k - \underline{A}^k}\right) A^{-1} dA = \left(\frac{k\tilde{A}^k}{\tilde{A}^k - 1}\right) \ln\tilde{A}$. ■

Proposition D.6 $\mu = \begin{cases} \left(\frac{k+2}{k}\right) \frac{\tilde{A}^2\left(\tilde{A}^k-1\right)-k\tilde{A}\left(\tilde{A}-1\right)}{\tilde{A}\left(\tilde{A}^{k+1}-1\right)-(k+1)\left(\tilde{A}-1\right)} & \text{, } \textit{if } k \neq 0, -1, -2 \\ \frac{(\tilde{A}-1)-\ln \tilde{A}}{\ln \tilde{A}-\left(\frac{\tilde{A}-1}{\tilde{A}}\right)} & \text{, } \textit{if } k = -1 \\ \left(\frac{1}{2}\right) \frac{(\tilde{A}-1)^2}{(\tilde{A}-1)-\ln \tilde{A}} & \text{, } \textit{if } k = -2 \end{cases}.$

Proof. If $k = -1$, using Proposition D.5 in Equation (14),

$$\mu = \frac{1 - \frac{\ln \tilde{A}}{\tilde{A}-1}}{\frac{\ln \tilde{A}}{\tilde{A}-1} - \left(\frac{\tilde{A}-1}{\tilde{A}^2-\tilde{A}}\right)} = \frac{(\tilde{A}-1) - \ln \tilde{A}}{\ln \tilde{A} - \left(\frac{\tilde{A}-1}{\tilde{A}}\right)}.$$

If $k = -2$, from Proposition D.5 and Equation (14),

$$\mu = \frac{1 - 2\left(\frac{\tilde{A}-1}{\tilde{A}^2-1}\right)}{2\left(\frac{\tilde{A}-1}{\tilde{A}^2-1}\right) - 2\left(\frac{\ln \tilde{A}}{\tilde{A}^2-1}\right)} = \left(\frac{1}{2}\right) \frac{(\tilde{A}^2-1) - 2(\tilde{A}-1)}{(\tilde{A}-1) - \ln \tilde{A}} = \left(\frac{1}{2}\right) \frac{(\tilde{A}-1)^2}{(\tilde{A}-1) - \ln \tilde{A}}.$$

Finally, if $k \neq 0, -1, -2$,

$$\mu = \left(\frac{k+2}{k}\right) \frac{(k+1) - k\left(\frac{\tilde{A}^{k+1}-1}{\tilde{A}^{k+1}-\tilde{A}}\right)}{(k+2)\left(\frac{\tilde{A}^{k+1}-1}{\tilde{A}^{k+1}-\tilde{A}}\right) - (k+1)\left(\frac{\tilde{A}^{k+2}-1}{\tilde{A}^{k+2}-\tilde{A}^2}\right)}$$

$$\mu = \left(\frac{k+2}{k}\right) \frac{(k+1)\left(\tilde{A}^{k+2} - \tilde{A}^2\right) - k\left(\tilde{A}^{k+2} - \tilde{A}\right)}{(k+2)\left(\tilde{A}^{k+2} - \tilde{A}\right) - (k+1)\left(\tilde{A}^{k+2} - 1\right)}$$

$$\mu = \left(\frac{k+2}{k}\right) \frac{\tilde{A}^{k+2} + k\tilde{A} - (k+1)\tilde{A}^2}{\tilde{A}^{k+2} + (k+1) - (k+2)\tilde{A}}$$

$$\mu = \left(\frac{k+2}{k}\right) \frac{\tilde{A}^2\left(\tilde{A}^k - 1\right) - k\tilde{A}\left(\tilde{A} - 1\right)}{\tilde{A}\left(\tilde{A}^{k+1} - 1\right) - (k+1)\left(\tilde{A} - 1\right)},$$

where we once again use Proposition D.5 in Equation (14). ■

Proposition D.7 $\lim_{\overline{A}\to+\infty} \mu = \begin{cases} \frac{k+2}{k} & \text{, } \textit{if } k > 0 \\ +\infty & \text{, } \textit{if } k < 0 \end{cases}.$

Proof. Note $\lim_{\overline{A}\to+\infty} \mu = \lim_{\tilde{A}\to+\infty} \mu$ as μ is only a function of $\tilde{A} \equiv \overline{A}/\underline{A}$ (Proposition D.6). Given that, it is sufficient to compute $\lim_{\tilde{A}\to+\infty} \mu$. If $k > 0$ and thus $k + j \neq 0$ for $j \in \mathbb{N} \setminus \{0\}$, from Proposition D.5 one gets

$$\lim_{\tilde{A}\to+\infty} \mathrm{E}_a\left((\underline{A}/A)^j\right) = \left(\frac{k}{k+j}\right) \lim_{\tilde{A}\to+\infty} \left(\frac{1 - \tilde{A}^{-k-j}}{1 - \tilde{A}^{-k}}\right) = \frac{k}{k+j},$$

which, from Equation (14), implies

$$\lim_{\tilde{A}\to+\infty}\mu=\lim_{\tilde{A}\to+\infty}\frac{1-\mathrm{E}_a\left(\underline{A}/A\right)}{\mathrm{E}_a\left(\underline{A}/A\right)-\mathrm{E}_a\left[(\underline{A}/A)^2\right]}=\frac{1-\frac{k}{k+1}}{\frac{k}{k+1}-\frac{k}{k+2}}=\frac{(k+2)\left[(k+1)-k\right]}{k(k+2)-k(k+1)}=\frac{k+2}{k}.$$

If $k<0$, using Proposition D.5 for $k+j\neq 0$ and $k+j=0$, respectively,

$$\lim_{\tilde{A}\to+\infty}\mathrm{E}_a\left((\underline{A}/A)^j\right)=\left(\frac{k}{k+j}\right)\lim_{\tilde{A}\to+\infty}\left(\frac{\tilde{A}^k-\tilde{A}^{-j}}{\tilde{A}^k-1}\right)=0$$

$$\lim_{\tilde{A}\to+\infty}\mathrm{E}_a\left((\underline{A}/A)^j\right)=\left(\lim_{\tilde{A}\to+\infty}\frac{k}{\tilde{A}^k-1}\right)\left(\lim_{\tilde{A}\to+\infty}\frac{\ln\tilde{A}}{\tilde{A}^{-k}}\right)=-k\left(\lim_{\tilde{A}\to+\infty}\frac{1/\tilde{A}}{-k\tilde{A}^{-k-1}}\right)=0,$$

where we apply L'Hôpital's rule in the last line, implying $\lim_{\tilde{A}\to+\infty}\mu=+\infty$ if $k<0$ from (14). ■

Proposition D.8 *For $k\neq 0$, $\tilde{A}\equiv\overline{A}/\underline{A}$, $\tilde{A}>1$, is a continuous, strictly increasing, and well-defined function of μ if and only if $\mu>1$ and $k<\frac{2}{\mu-1}$.*

Proof. Let $\hat{g}$ be the density of a truncated Pareto distribution with shape parameter $k\neq 0$ defined for $A\in[\underline{A},A_h]$, $A_h>\overline{A}>\underline{A}$, with $\hat{G}$ being the respective cumulative distribution function. It is easy to see $\tilde{g}(A)=\hat{g}(A)/\hat{G}(\overline{A})$ is the density of a truncated Pareto distribution with the same parameter $k\neq 0$ over the support $\left[\underline{A},\overline{A}\right]$. As a consequence, we can use Proposition D.4 and thus $\overline{A}$, $\overline{A}>\underline{A}$, is continuous, strictly increasing, and well defined in μ if and only if $\mu\in(1,\lim_{\overline{A}\to+\infty}\mu)$. From Proposition D.6, μ is only a function of $\tilde{A}\equiv\overline{A}/\underline{A}$ (given k), implying these features of $\overline{A}$ also hold for $\tilde{A}$. Moreover, given Proposition D.7, one can rewrite the condition $\mu\in(1,\lim_{\overline{A}\to+\infty}\mu)$ as (i) $\mu>1$ and $k<0$ or (ii) $1<\mu<\frac{k+2}{k}$ and $k>0$. Note $1<\mu<\frac{k+2}{k}\to k<\frac{2}{\mu-1}$, which is always fulfilled for $k<0$ given $\mu>1$. Therefore, for any $k\neq 0$, $\mu\in(1,\lim_{\overline{A}\to+\infty}\mu)$ holds if and only if $\mu>1$ and $k<\frac{2}{\mu-1}$. ■

Figure D.1 illustrates the results of Proposition D.8, plotting $\tilde{A}$ against μ for truncated Pareto distributions with $k=3,5,9$ and a uniform distribution ($k=-1$).

From Proposition D.8, given $\mu>1$, a solution for the calibration algorithm exists for any strictly negative shape parameter k. Hence, it is feasible even for highly negative values of k. However, an important inquiry arises regarding the behavior of the estimated Ω under such extreme conditions. This question is addressed in the concluding proposition of this section:

Proposition D.9 $\lim_{k\to-\infty}\Omega=1$.

Proof. Initially, use Proposition D.5 to get that, for $j\in\mathbb{N}\setminus\{0\}$,

$$\lim_{k\to-\infty}\mathrm{E}_a\left((\underline{A}/A)^j\right)=\left[\lim_{k\to-\infty}\left(\frac{k}{k+j}\right)\right]\left[\lim_{k\to-\infty}\left(\frac{\tilde{A}^{k+j}-1}{\tilde{A}^{k+j}-\tilde{A}^j}\right)\right]=1\times\left(\frac{0-1}{0-\tilde{A}^j}\right)=\tilde{A}^{-j}.$$

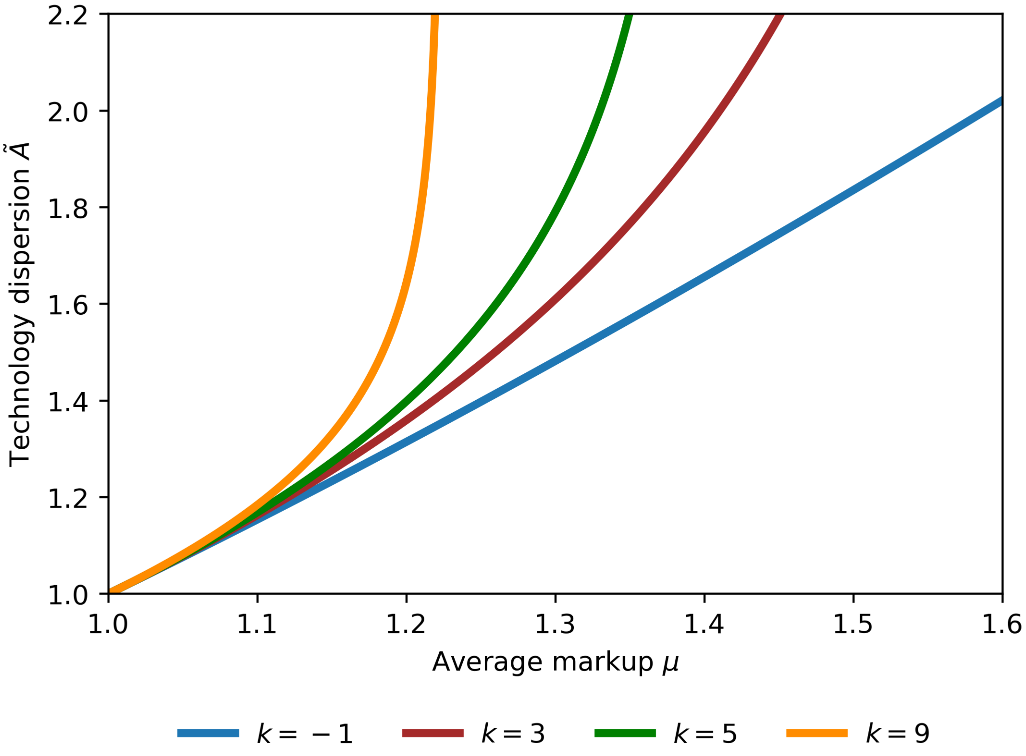

Figure D.1: Technological dispersion $\tilde{A} \equiv \overline{A}/\underline{A}$ *versus* average markup μ.

As a result, from Equation (11) holding exactly,

$$\lim_{k\to-\infty} \Omega = \lim_{k\to-\infty} \frac{\tilde{A}^{-1}\left[1 - \mathrm{E}_a\left(\underline{A}/A\right)\right]}{\mathrm{E}_a\left(\underline{A}/A\right) - \mathrm{E}_a\left[(\underline{A}/A)^2\right]} = \frac{\tilde{A}^{-1}\left(1 - \tilde{A}^{-1}\right)}{\tilde{A}^{-1} - \tilde{A}^{-2}} = \frac{\tilde{A}-1}{\tilde{A}-1} = 1,$$

where we use $\lim_{k\to-\infty} \mathrm{E}_a\left((\underline{A}/A)^j\right) = \tilde{A}^{-j}$ for $j = 1, 2$. ■

E Robustness of misallocation estimates to markup level

We want to prove that, for $k = -1$ and $\mu > 1$, estimates of allocative efficiency *growth* are highly robust to (i) the level of LS and (ii) the choice of α. More precisely, let us show $\Delta \ln \Omega$ is highly robust to the level of $\mu = \frac{1-\alpha}{LS}$. Initially, use Equation (12) to get

$$\Omega = \mu/\tilde{A} \rightarrow \Delta \ln \Omega = \Delta \ln \mu - \Delta \ln \tilde{A}, \tag{E.1}$$

implying $\Delta \ln \Omega$ is independent of the level of μ if $\ln \tilde{A} = \gamma_0 + \gamma_1 \ln \mu$, when (E.1) becomes $\Delta \ln \Omega = (1 - \gamma_1)\Delta \ln \mu$. We use this idea to find the value of k that results in $\Delta \ln \Omega$ being nearly independent of the level of μ. Specifically, we identify the k that provides the best fit for the regression $\ln \tilde{A}(\mu; k) = \gamma_0 + \gamma_1 \ln \mu + \varepsilon$, where $\tilde{A}(\mu; k)$ is implicitly defined in (14). We consider 100 observations for μ between 1 and μ_h, for μ_h equal to 1.05, 1.15, 1.3, 1.5, 3, 5 or 10.[37] For each μ_h and thus each set of observations of μ, choosing a k, we obtain the respective 100 observations of $\tilde{A}$ numerically from (14), estimating γ_0 and γ_1 using Ordinary Least Squares (OLS). Hence, for each μ_h, we can obtain the sum of squared residuals (SSR) associated with any $k < \frac{2}{\mu_h - 1}$, $k \neq 0$, allowing the numerical search of the robust k that minimizes the SSR.

The robust k, denoted by k^*, and the associated (centered) R^2 for each μ_h are shown respectively in the second and third columns of Table E.1. As can be seen, the R^2 is always very high, with k^* being close to -1, particularly for smaller μ_h. In any case, even when k^* is not so close to -1, the R^2 associated with $k = -1$ is still very high (fourth column of Table E.1), indicating the estimate of $\Delta \ln \Omega$ under $k = -1$ is also highly robust to the level of μ.

Another way to see that is evaluating $\frac{\partial \ln \tilde{A}(\mu; -1)}{\partial \ln \mu}$. Given μ_h and the correspondent set of observations $\{\mu_i\}$, it is possible to show $\frac{\partial \ln \tilde{A}(\mu_i; -1)}{\partial \ln \mu} \in \left[\frac{\partial \ln \tilde{A}(\mu_h; -1)}{\partial \ln \mu}, 1.5\right) \subset (1, 1.5)$. These derivative ranges for each considered μ_h are shown in the last column of Table E.1.[38] The ranges are always very narrow, especially for low μ_h, confirming the estimate of $\Delta \ln \Omega$ under $k = -1$ is highly robust to the choice of α and the level of LS.[39]

[37] Empirical evidence from firm-level data suggests $\mu = 10$ is high enough (see De Loecker et al. 2020 for 1955–2016 US, Traina 2018 for 1950–2016 US, Baqaee and Farhi 2020 for 1985–2015 US, and De Loecker and Eeckhout 2018 for 134 countries between 1980 and 2016). For instance, De Loecker and Eeckhout (2018) find global cost-weighted average markups are always below 1.6. The sales-weighted average, which is typically a higher measure (De Loecker et al. 2020), remains below 2 for global regions (Europe, North America, South America, Asia, Oceania, and Africa) and under 4 for a selection of 40 countries shown in their Appendix A.

[38] To get $\frac{\partial \ln \tilde{A}(\mu_h; -1)}{\partial \ln \mu}$, we first obtain $\tilde{A}(\mu_h; -1)$ numerically from (E.2) and then compute (E.3) to get $\frac{\partial \ln \mu\left(\tilde{A}(\mu_h; -1); -1\right)}{\partial \ln \tilde{A}}$. Finally, we use Proposition E.4 shown below and compute $\frac{\partial \ln \tilde{A}(\mu_h; -1)}{\partial \ln \mu} = \left[\frac{\partial \ln \mu\left(\tilde{A}(\mu_h; -1); -1\right)}{\partial \ln \tilde{A}}\right]^{-1}$.

[39] Up to a first-order Taylor approximation around $\mu^* > 1$, $\ln \tilde{A}(\mu; -1) \approx \ln \tilde{A}(\mu^*; -1) + \frac{\partial \ln \tilde{A}(\mu^*; -1)}{\partial \ln \mu}(\ln \mu - \ln \mu^*)$. However, as $\mu^* \to 1^+$, $\tilde{A}(\mu; -1) \to 1$ (Propositions D.3 and D.4) and $\frac{\partial \ln \tilde{A}(\mu^*; -1)}{\partial \ln \mu} \to 1.5$ (Proposition E.4 shown below). Consequently, for low average markup, $\ln \tilde{A}(\mu; -1) \approx 1.5 \ln \mu \to \tilde{A}(\mu; -1) \approx \mu^{1.5}$ and thus, from (E.1), $\Delta \ln \Omega \approx -0.5 \Delta \ln \mu = 0.5 \Delta \ln LS$.

Table E.1: Robustness evaluation for different μ_h

μ_h	$k=k^*$		$k=-1$	
	k^*	R^2	R^2	$\frac{\partial \ln \tilde{A}}{\partial \ln \mu}$ range
1.05	-0.9949	1.000000	1.000000	[1.4999,1.5)
1.15	-0.9860	1.000000	1.000000	[1.4993,1.5)
1.3	-0.9738	1.000000	1.000000	[1.4974,1.5)
1.5	-0.9596	1.000000	1.000000	[1.4939,1.5)
3	-0.8931	1.000000	0.999986	[1.4590,1.5)
5	-0.8480	0.999998	0.999944	[1.4205,1.5)
10	-0.7940	0.999994	0.999817	[1.3632,1.5)

This derivative assessment relied on some results. We prove them now. To start, note μ is a function of $\tilde{A}$ from Equation (14), which for the uniform distribution is equivalent to

$$\mu\left(\tilde{A};-1\right)=\frac{\left(\tilde{A}-1\right)-\ln\tilde{A}}{\ln\tilde{A}-\left(\frac{\tilde{A}-1}{\tilde{A}}\right)} \tag{E.2}$$

due to Proposition D.6. As a consequence,

$$\begin{aligned}\frac{\partial\ln\mu\left(\tilde{A};-1\right)}{\partial\ln\tilde{A}}&=\frac{\tilde{A}-1}{\left(\tilde{A}-1\right)-\ln\tilde{A}}-\frac{1-\frac{\tilde{A}}{\tilde{A}^2}}{\ln\tilde{A}-\left(\frac{\tilde{A}-1}{\tilde{A}}\right)}=\frac{\tilde{A}-1}{\left(\tilde{A}-1\right)-\ln\tilde{A}}-\frac{\tilde{A}-1}{\tilde{A}\ln\tilde{A}-\left(\tilde{A}-1\right)}\\ \frac{\partial\ln\mu\left(\tilde{A};-1\right)}{\partial\ln\tilde{A}}&=\frac{\left(\tilde{A}-1\right)\left[\left(\tilde{A}+1\right)\ln\tilde{A}-2\left(\tilde{A}-1\right)\right]}{\left[\left(\tilde{A}-1\right)-\ln\tilde{A}\right]\left[\tilde{A}\ln\tilde{A}-\left(\tilde{A}-1\right)\right]}\end{aligned} \tag{E.3}$$

$$\begin{aligned}\frac{\partial^2\ln\mu\left(\tilde{A};-1\right)}{\partial^2\ln\tilde{A}}&=\frac{\tilde{A}\left[\left(\tilde{A}-1\right)-\ln\tilde{A}\right]-\left(\tilde{A}-1\right)^2}{\left[\left(\tilde{A}-1\right)-\ln\tilde{A}\right]^2}-\frac{\tilde{A}\left[\tilde{A}\ln\tilde{A}-\left(\tilde{A}-1\right)\right]-\left(\tilde{A}-1\right)\tilde{A}\ln\tilde{A}}{\left[\tilde{A}\ln\tilde{A}-\left(\tilde{A}-1\right)\right]^2}\\ \frac{\partial^2\ln\mu\left(\tilde{A};-1\right)}{\partial^2\ln\tilde{A}}&=\frac{\left(\tilde{A}-1\right)-\tilde{A}\ln\tilde{A}}{\left[\left(\tilde{A}-1\right)-\ln\tilde{A}\right]^2}-\frac{\tilde{A}\left[\ln\tilde{A}-\left(\tilde{A}-1\right)\right]}{\left[\tilde{A}\ln\tilde{A}-\left(\tilde{A}-1\right)\right]^2}\\ \frac{\partial^2\ln\mu\left(\tilde{A};-1\right)}{\partial^2\ln\tilde{A}}&=\frac{\tilde{A}\left[\left(\tilde{A}-1\right)-\ln\tilde{A}\right]^3-\left[\tilde{A}\ln\tilde{A}-\left(\tilde{A}-1\right)\right]^3}{\left[\left(\tilde{A}-1\right)-\ln\tilde{A}\right]^2\left[\tilde{A}\ln\tilde{A}-\left(\tilde{A}-1\right)\right]^2}.\end{aligned} \tag{E.4}$$

Proposition E.1 *For* $\tilde{A}>1$, $\frac{\partial^2\ln\mu\left(\tilde{A};-1\right)}{\partial^2\ln\tilde{A}}>0$.

Proof. Let $\tilde{A}>1$. In this case, (i) for $g(\tilde{A})\equiv\left(\tilde{A}-1\right)-\ln\tilde{A}$, $g'(\tilde{A})=\frac{\tilde{A}-1}{\tilde{A}}>0\xrightarrow{g(1)=0}$

$g(\tilde{A}) > 0$, and (ii) for $h(\tilde{A}) \equiv \tilde{A}\ln\tilde{A} - \left(\tilde{A}-1\right)$, $h'(\tilde{A}) = \ln\tilde{A} > 0 \xrightarrow{h(1)=0} h(\tilde{A}) > 0$. Consequently, the denominator of (E.4) is strictly positive. Hence, it is sufficient to show $f(\tilde{A}) \equiv \sqrt[3]{\tilde{A}}\left[(\tilde{A}-1)-\ln\tilde{A}\right] - \left[\tilde{A}\ln\tilde{A}-(\tilde{A}-1)\right] > 0$. Note $f(1)=0$ and

$$
\begin{aligned}
f'(\tilde{A}) =& \frac{\sqrt[3]{\tilde{A}}}{3\tilde{A}}\left[(\tilde{A}-1)-\ln\tilde{A}\right] + \sqrt[3]{\tilde{A}}\left(\frac{\tilde{A}-1}{\tilde{A}}\right) - \ln\tilde{A} \\
f'(\tilde{A}) =& \frac{\sqrt[3]{\tilde{A}}}{3\tilde{A}}\left[4(\tilde{A}-1)-\ln\tilde{A}\right] - \ln\tilde{A} \to f'(1) = 0
\end{aligned}
$$

$$
\begin{aligned}
f''(\tilde{A}) =& -2\frac{\sqrt[3]{\tilde{A}}}{9\tilde{A}^2}\left[4(\tilde{A}-1)-\ln\tilde{A}\right] + \frac{\sqrt[3]{\tilde{A}}}{3\tilde{A}}\left(\frac{4\tilde{A}-1}{\tilde{A}}\right) - \frac{1}{\tilde{A}} \\
f''(\tilde{A}) =& \frac{\sqrt[3]{\tilde{A}}}{9\tilde{A}^2}\left[-8(\tilde{A}-1)+2\ln\tilde{A}+12\tilde{A}-3-9\tilde{A}^{2/3}\right] = \frac{\sqrt[3]{\tilde{A}}}{9\tilde{A}^2}\left[4\tilde{A}+5+2\ln\tilde{A}-9\tilde{A}^{2/3}\right]
\end{aligned}
$$

$$
\begin{aligned}
\tilde{f}(\tilde{A}) \equiv& 4\tilde{A}+5+2\ln\tilde{A}-9\tilde{A}^{2/3} \to \tilde{f}(1) = 0 \\
\tilde{f}'(\tilde{A}) =& 4 + \frac{2}{\tilde{A}} - \frac{6}{\sqrt[3]{\tilde{A}}} \to \tilde{f}'(1) = 0 \\
\tilde{f}''(\tilde{A}) =& -\frac{2}{\tilde{A}^2} + \frac{2}{\tilde{A}\sqrt[3]{\tilde{A}}} = \frac{2}{\tilde{A}^2}\left(\tilde{A}^{2/3}-1\right) \to \tilde{f}''(\tilde{A}) > 0 \text{ for } \tilde{A} > 1.
\end{aligned}
$$

Therefore, for $\tilde{A} > 1$, $\tilde{f}''(\tilde{A}) > 0 \xrightarrow{\tilde{f}'(1)=0} \tilde{f}'(\tilde{A}) > 0 \xrightarrow{\tilde{f}(1)=0} \tilde{f}(\tilde{A}) > 0 \to f''(\tilde{A}) > 0 \xrightarrow{f'(1)=0} f'(\tilde{A}) > 0 \xrightarrow{f(1)=0} f(\tilde{A}) > 0$. ■

Proposition E.2 $\lim_{\tilde{A}\to+\infty} \frac{\partial \ln\mu\left(\tilde{A};-1\right)}{\partial\ln\tilde{A}} = 1$.

Proof. From Equation (E.3),

$$
\begin{aligned}
\lim_{\tilde{A}\to+\infty} \frac{\partial\ln\mu\left(\tilde{A};-1\right)}{\partial\ln\tilde{A}} =& \frac{1}{1-\lim_{\tilde{A}\to+\infty}\frac{\ln\tilde{A}}{\tilde{A}-1}} - \frac{1}{\lim_{\tilde{A}\to+\infty}\frac{\tilde{A}\ln\tilde{A}}{\tilde{A}-1}-1} \\
\lim_{\tilde{A}\to+\infty} \frac{\partial\ln\mu\left(\tilde{A};-1\right)}{\partial\ln\tilde{A}} =& \frac{1}{1-\lim_{\tilde{A}\to+\infty}\frac{1}{\tilde{A}}} - \frac{1}{\lim_{\tilde{A}\to+\infty}\left(\ln\tilde{A}+1\right)-1} = 1-0 = 1,
\end{aligned}
$$

where we apply L'Hôpital's rule in the second line. ■

Proposition E.3 $\lim_{\tilde{A}\to1^+} \frac{\partial \ln\mu\left(\tilde{A};-1\right)}{\partial\ln\tilde{A}} = \frac{2}{3}$.

Proof. From Equation (E.3),

$$\lim_{\tilde{A}\to 1^+} \frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} = \lim_{\tilde{A}\to 1^+} \frac{\left[\left(\tilde{A}+1\right)\ln \tilde{A} - 2\left(\tilde{A}-1\right)\right] + \left(\tilde{A}-1\right)\left(\ln \tilde{A} + \frac{\tilde{A}+1}{\tilde{A}} - 2\right)}{\left(\frac{\tilde{A}-1}{\tilde{A}}\right)\left[\tilde{A}\ln \tilde{A} - \left(\tilde{A}-1\right)\right] + \left[\left(\tilde{A}-1\right) - \ln \tilde{A}\right]\ln \tilde{A}}$$

$$\lim_{\tilde{A}\to 1^+} \frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} = \lim_{\tilde{A}\to 1^+} \frac{\left(\ln \tilde{A} + \frac{\tilde{A}+1}{\tilde{A}} - 2\right) + \left(\ln \tilde{A} + \frac{\tilde{A}+1}{\tilde{A}} - 2\right) + \left(\tilde{A}-1\right)\left(\frac{1}{\tilde{A}} - \frac{1}{\tilde{A}^2}\right)}{\left(\frac{1}{\tilde{A}^2}\right)\left[\tilde{A}\ln \tilde{A} - \left(\tilde{A}-1\right)\right] + \left(\frac{\tilde{A}-1}{\tilde{A}}\right)\ln \tilde{A} + \left(\frac{\tilde{A}-1}{\tilde{A}}\right)\ln \tilde{A} + \left(\frac{\tilde{A}-1}{\tilde{A}} - \frac{\ln \tilde{A}}{\tilde{A}}\right)}$$

$$\lim_{\tilde{A}\to 1^+} \frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} = \lim_{\tilde{A}\to 1^+} \tilde{A}\frac{2\left(\ln \tilde{A} + \frac{\tilde{A}+1}{\tilde{A}} - 2\right) + \left(\frac{\tilde{A}-1}{\tilde{A}}\right)^2}{\ln \tilde{A} - \frac{\tilde{A}-1}{\tilde{A}} + 2\left(\tilde{A}-1\right)\ln \tilde{A} + \left(\tilde{A}-1\right) - \ln \tilde{A}}$$

$$\lim_{\tilde{A}\to 1^+} \frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} = \lim_{\tilde{A}\to 1^+} \frac{2\left(\frac{1}{\tilde{A}} - \frac{1}{\tilde{A}^2}\right) + 2\left(\frac{\tilde{A}-1}{\tilde{A}}\right)\left(\frac{1}{\tilde{A}^2}\right)}{-\frac{1}{\tilde{A}^2} + 2\ln \tilde{A} + 2\left(\frac{\tilde{A}-1}{\tilde{A}}\right) + 1}$$

$$\lim_{\tilde{A}\to 1^+} \frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} = 2\lim_{\tilde{A}\to 1^+} \frac{\left(\frac{\tilde{A}-1}{\tilde{A}^2}\right)\left(\frac{\tilde{A}+1}{\tilde{A}}\right)}{2\ln \tilde{A} + 3 - \frac{2}{\tilde{A}} - \frac{1}{\tilde{A}^2}} = 2\lim_{\tilde{A}\to 1^+} \frac{\tilde{A}^2 - 1}{2\tilde{A}^3\ln \tilde{A} + 3\tilde{A}^3 - 2\tilde{A}^2 - \tilde{A}}$$

$$\lim_{\tilde{A}\to 1^+} \frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} = 2\lim_{\tilde{A}\to 1^+} \frac{2\tilde{A}}{6\tilde{A}^2\ln \tilde{A} + 2\tilde{A}^2 + 9\tilde{A}^2 - 4\tilde{A} - 1} = 2\left(\frac{2}{6}\right) = \frac{2}{3},$$

where we apply L'Hôpital's rule in the first, second, fourth, and sixth lines. ■

Proposition E.4 $\frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu} = \left[\frac{\partial \ln \mu\left(\tilde{A}(\mu;-1);-1\right)}{\partial \ln \tilde{A}}\right]^{-1}$ *is continuous and strictly decreasing in* $\mu > 1$, *with* $\lim_{\mu\to+\infty} \frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu} = 1$ *and* $\lim_{\mu\to 1^+} \frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu} = 1.5$.

Proof. From Proposition D.8, $\tilde{A}(\mu;-1)$, which is implicitly defined in (E.2), is continuous, strictly increasing, and well defined for $\mu > 1$. Moreover,

$$1 = \frac{\partial \ln \tilde{A}}{\partial \ln \tilde{A}} = \frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu}\frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} \to \frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu} = \left[\frac{\partial \ln \mu\left(\tilde{A}(\mu;-1);-1\right)}{\partial \ln \tilde{A}}\right]^{-1}.$$

Hence, since $\frac{\partial \ln \mu(\tilde{A};-1)}{\partial \ln \tilde{A}}$ is continuous and strictly increasing in $\tilde{A} > 1$ (Proposition E.1), $\frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu}$ is continuous and strictly decreasing in $\mu > 1$. Furthermore, using Propositions E.2 and E.3,

$$\lim_{\mu\to+\infty} \frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu} = \lim_{\tilde{A}\to+\infty} \frac{1}{\frac{\partial \ln \mu(\tilde{A};-1)}{\partial \ln \tilde{A}}} = 1$$

$$\lim_{\mu\to 1^+} \frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu} = \lim_{\tilde{A}\to 1^+} \frac{1}{\frac{\partial \ln \mu(\tilde{A};-1)}{\partial \ln \tilde{A}}} = 1.5,$$

since (i) $\tilde{A}(\mu; -1) \to +\infty$ if and only if $\mu \to +\infty$ (Propositions D.7 and D.8) and (ii) $\tilde{A}(\mu; -1) \to 1^+$ if and only if $\mu \to 1^+$ (Propositions D.3 and D.4). ■

F Model extensions

We develop two model extensions. In both cases, we consider a discrete number of firms as in Section 2.1, but they can easily be adapted to a continuum of firms following the approach of Section 2.2.

F.1 Beyond Cobb-Douglas production functions

In this section, we generalize the model by considering an arbitrary well-behaved production function with M factors of production, provided it exhibits (i) constant returns to scale and (ii) Hicks-neutral productivity shifter. Formally, the production of each firm $i \in \{1, 2, ..., N\}$ is now given by

$$Y_i = A_i f(F_{1i}, ..., F_{Mi}), \tag{2'}$$

where $A_i > 0$ and $F_{ji} \geq 0$ is the quantity of factor j used by firm i. The function f is (i) homogeneous of degree 1, (ii) differentiable, and (iii) well-behaved in the sense that the firms' cost minimization problem has a unique interior solution for strictly greater than zero rental prices.

Before discussing the derivation of the model, note if each firm i uses its inputs optimally, taking the same inputs' rental prices as given, the FOC of its cost minimization problem implies that, for any inputs r and k,

$$\frac{w_r}{MPF_{ri}} = \frac{w_k}{MPF_{ki}} \leftrightarrow \frac{w_r}{w_k} = \frac{A_i f_r(F_{1i}, ..., F_{Mi})}{A_i f_k(F_{1i}, ..., F_{Mi})} = \frac{f_r(1, F_{2i}/F_{1i}, ..., F_{Mi}/F_{1i})}{f_k(1, F_{2i}/F_{1i}, ..., F_{Mi}/F_{1i})}, \tag{A.3'}$$

where $w_r > 0$ is the rental price of factor r, MPF_{ri} is firm i marginal product of factor r, $f_r(F_{1i}, ..., F_{Mi}) \equiv \frac{\partial f(F_{1i}, ..., F_{Mi})}{\partial F_{ri}}$, and in the last part we use that f_r is homogeneous of degree 0 as f is homogeneous of degree 1. Setting $r = 1, 2, ..., k-1, k+1, ..., M$ on (A.3'), we would get a system of $M-1$ equations in the $M-1$ unknowns $F_{2i}/F_{1i}, ..., F_{Mi}/F_{1i}$. Since the cost minimization problem has a unique interior solution by assumption, the FOCs are satisfied in the optimal, and thus, there is a unique solution for this system. Hence, since all firms face the same problem, they choose the same relative quantities of factors, that is, $F_{ri}/F_{ki} = F_{rj}/F_{kj}$ for any firms i and j and factors r and k.

F.1.1 Derivation of the model

Given its similarity to the baseline model, we do not discuss the derivation in detail, emphasizing solely the main differences from the Cobb-Douglas case.

Environment and technology. The only difference is the firms' production function, which is now represented by (2') instead of (2). Aggregate output Y continues to be given by (1).

Market competition and optimal decision. As the marginal cost remains invariant to output, Equations (3) and (4) of the Cobb-Douglas case persist, albeit with a distinct marginal cost function. The SOC is the same (A.1) and thus holds.

Equilibrium allocation. Since firms use inputs optimally, $F_{ri}/F_{ki} = F_{rj}/F_{kj}$ and $MC_i = \frac{w_k}{MPF_{ki}} = \frac{w_k}{A_i f_k(1, F_{2i}/F_{1i}, ..., F_{Mi}/F_{1i})}$ for any firms i and j and factors r and k. As a consequence, $\frac{MC_i}{MC_j} = \frac{A_j}{A_i}$ as before, implying Equations (5), (A.2), (6), and (7) are still valid.

Aggregate productivity and misallocation. Naturally, the aggregate production function is not (8) anymore. However, using (1) and (2'), one easily obtains the new expression:

$$Y = \sum_{i=1}^{N} A_i f(F_{1i}, ..., F_{Mi}) = \overline{A}\Omega f(F_1, ..., F_M), \tag{8'}$$

where $F_j \equiv \sum_{i=1}^{N} F_{ji}$, and $\Omega \equiv \sum_{i=1}^{N} \frac{f(F_{1i},...,F_{Mi})}{f(F_1,...,F_M)} \left(A_i/\overline{A}\right)$.

Since each firm uses its inputs optimally, taking the same inputs' rental prices as given, every active firm chooses the same relative quantities of factors, implying

$$\theta_{rj} \equiv \frac{F_{rj}}{\sum_{i=1}^{N} F_{ri}} = \frac{F_{kj}\frac{F_{rj}}{F_{kj}}}{\sum_{i=1}^{N} F_{ki}\frac{F_{ri}}{F_{ki}}} = \frac{F_{kj}\frac{F_{rj}}{F_{kj}}}{\sum_{i=1}^{N} F_{ki}\frac{F_{rj}}{F_{kj}}} = \frac{F_{kj}}{\sum_{i=1}^{N} F_{ki}} \equiv \theta_{kj}, \tag{A.4'}$$

which allows us to define $\theta_j \equiv \theta_{1j} = \theta_{2j} = ... = \theta_{Mj}$ for each firm j and factors r and k. As a result, from Equations (2') and (8'),

$$s(A_i) = \frac{Y_i}{Y} = \frac{A_i f(F_{1i}, ..., F_{Mi})}{\overline{A}\Omega f(F_1, ..., F_M)} = \frac{A_i f(\theta_i F_1, ..., \theta_i F_M)}{\overline{A}\Omega f(F_1, ..., F_M)} = \frac{A_i \theta_i}{\overline{A}\Omega} \rightarrow \theta_i = \overline{A}\Omega \frac{s(A_i)}{A_i}, \tag{9'}$$

which is exactly equal to (9). Consequently, the aggregate TFP $\overline{A}\Omega$ continue to be the quantity-weighted harmonic mean of firms' productivity shown in (10), with allocative efficiency Ω still given by Equation (11) and thus showing the same properties as before.

Average markup. Given that the main results from the baseline model continue to be valid, it is easy to see that cost-weighted average of firm-level markups $\mu \equiv \sum_{i=1}^{N} \left(\frac{\sum_{k=1}^{M} F_{ki} w_k}{\sum_{k=1}^{M} F_k w_k} \right) \mu_i = \sum_{i=1}^{N} \theta_i \mu_i$ is still given by Equation (12).

F.1.2 Quantification strategy

The target moments $\overline{A}\Omega$ and μ can be computed in the model using the same expressions of the Cobb-Douglas case. Consequently, given data on such moments, we can empirically implement this generalized model following exactly the strategy described in Section 3.3. The difference between models appears only when computing those target moments from data.

Analogously to the baseline case, we get $\overline{A}\Omega$ and μ using standard macroeconomic data and the parameters of the production function. On the one hand, TFP continues to be backed out as a residual in the aggregate output function, which is now given by Equation (8'): $\overline{A}\Omega = \frac{Y}{f(F_1,...,F_M)}$. On the other hand, note for any firm i, $\mu_i = \frac{Y_i p}{H_i w}\frac{H_i MPH_i}{Y_i} = \frac{\alpha_H}{LS_i}$, where $\alpha_H \equiv \frac{H_i MPH_i}{Y_i} = \frac{H_i f_1(F_{1i},F_{2i},...,F_{Mi})}{f(F_{1i},F_{2i},...,F_{Mi})} = \frac{f_1(1,F_{2i}/H_i,...,F_{Mi}/H_i)}{f(1,F_{2i}/H_i,...,F_{Mi}/H_i)} = \frac{f_1(1,F_2/H,...,F_M/H)}{f(1,F_2/H,...,F_M/H)}$ is the elasticity of the production function to human capital, which we assume is factor 1. As a result, $\mu = \sum_{i=1}^N \theta_i\mu_i = \sum_{i=1}^N \left(\frac{H_i}{H}\right)\left[\alpha_H \frac{Y_i p}{H_i w}\right] = \alpha_H \frac{p\sum_{i=1}^N Y_i}{Hw} = \frac{\alpha_H}{LS}$. Note computing $\alpha_H = \frac{f_1(1,F_2/H,...,F_M/H)}{f(1,F_2/H,...,F_M/H)}$ requires, in general, the relative aggregate quantities of factors, while in the special Cobb-Douglas case it does not as α_H would be simply the share parameter associated with human capital. Indeed, for the baseline model, in which $f(H,K) = K^\alpha H^{1-\alpha}$, $\alpha_H = \frac{f_1(1,K/H)}{f(1,K/H)} = \frac{(1-\alpha)(K/H)^\alpha(H/H)^{-\alpha}}{(H/H)^\alpha(H/H)^{1-\alpha}} = 1-\alpha$, implying $\mu = \frac{1-\alpha}{LS}$, which is exactly the result shown in Section 3.2. Thus, differently from the Cobb-Douglas case, α_H is not generally time invariant even if the production function parameters are kept constant, meaning the average markup μ growth may differ from the labor share growth.

F.1.3 Calibration of the production function

In Section 4.1, we calibrate the baseline Cobb-Douglas production function for the US using cost share data. A similar approach can be employed here, since

$$\begin{aligned}
\frac{Hw}{\sum_{k=1}^M F_k w_k} &= \frac{\theta_i Hw}{\sum_{k=1}^M \theta_i F_k w_k} = \frac{H_i}{\sum_{k=1}^M F_{ki}\frac{w_k}{w}} = \frac{H_i}{\sum_{k=1}^M F_{ki}\frac{f_k(1,F_{2i}/H_i,...,F_{Mi}/H_i)}{f_1(1,F_{2i}/H_i,...,F_{Mi}/H_i)}} \\
\frac{Hw}{\sum_{k=1}^M F_k w_k} &= \frac{f_1(1,F_{2i}/H_i,...,F_{Mi}/H_i)}{\sum_{k=1}^M \frac{F_{ki}}{H_i} f_k(1,F_{2i}/H_i,...,F_{Mi}/H_i)} = \frac{f_1(1,F_2/H,...,F_M/H)}{f(1,F_2/H,...,F_M/H)} = \alpha_H, \quad \text{(A.5')}
\end{aligned}$$

where we use, in the first line, (A.3') with $F_{1i} = H_i$ and, in the last line, the Euler's Theorem for Homogeneous Functions and that, in the optimal, $F_{ki}/H_i = F_k/H$ for any factor k.[40] As a result, one can calibrate the production function by finding the parameters of f that make $\alpha_H = \frac{f_1(1,F_2/H,...,F_M/H)}{f(1,F_2/H,...,F_M/H)}$ closest (in some sense) to labor cost share data. This calibration could also include cost share data for other factors of production, using $\alpha_r = \frac{f_r(F_1/F_r,...,F_M/F_r)}{f(F_1/F_r,...,F_M/F_r)} = \frac{F_r w_r}{\sum_{k=1}^M F_k w_k}$ for any factor r. However, one may use at most $M-1$ factors' cost share data in the calibration, since share data from the remaining factor would not be informative as $\sum_{r=1}^M \alpha_r = 1$.

To illustrate this calibration procedure, let us consider the CES production function

$$Y_i = A_i\left[\alpha K_i^{\frac{\sigma-1}{\sigma}} + (1-\alpha) H_i^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \quad \text{(F.1)}$$

[40]As expected, Equation (A.5') is consistent with (A.5), since in the baseline case $1-\alpha = \alpha_H = \frac{Hw}{Kr+Hw}$.

where $\sigma > 0$ is the elasticity of substitution and $\alpha \in (0,1)$. In this case,

$$\alpha_H = \frac{f_1(1,K/H)}{f(1,K/H)} = \frac{f_1(H,K)}{f(1,K/H)} = \frac{\left(\frac{\sigma}{\sigma-1}\right)\left[\alpha K^{\frac{\sigma-1}{\sigma}} + (1-\alpha) H^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}} (1-\alpha)\left(\frac{\sigma-1}{\sigma}\right) H^{\frac{-(\sigma-1)/\sigma}{\sigma-1}}}{\left[\alpha (K/H)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\right]^{\frac{\sigma}{\sigma-1}}}$$

$$\alpha_H = \frac{(1-\alpha)\left[\alpha (K/H)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\right]^{\frac{1}{\sigma-1}}}{\left[\alpha (K/H)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\right]^{\frac{\sigma}{\sigma-1}}} = \frac{1-\alpha}{\alpha (K/H)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)}, \tag{F.2}$$

with $\alpha_H \to 1-\alpha$ as $\sigma \to 1$, which is expected as the CES function (F.1) converges to the baseline Cobb-Douglas function (2) in this limit case. Hence, given data on K and H and the parameters α and σ, we can compute α_H from (F.2), allowing us to gauge α and σ by matching α_H to labor cost share data.

F.2 Firm-specific wedges

We consider the same setup of the baseline model but add one wedge for each potential entrant firm, including a firm-specific tax rate over revenue $\tau_i = \tau(A_i) \in (-\infty, 1)$.

F.2.1 Derivation of the model

As in Appendix F.1.1, we do not provide a complete derivation. Instead, we just highlight the key differences from the baseline model.

Environment and technology. The environment and technology remain unchanged, with Equations (1) and (2) still holding.

Market competition and optimal decision. Since now there is a revenue tax, each firm $i \in \{1, 2, ..., N\}$ solves a slightly different profit maximization problem:

$$\begin{aligned} \max_{Y_i} \quad & p\,(1-\tau(A_i))\,Y_i - wH_i - rK_i = \max_{Y_i} \quad [p\,(1-\tau(A_i)) - MC_i]\,Y_i \\ \text{s.t.} \quad & w > 0, r > 0, p = p(Y), Y_j \geq 0\ \forall j \in \{1, 2, ..., N\} \setminus \{i\} \end{aligned} . \tag{3''}$$

The FOC of this optimization problem is

$$p\,(1-\tau(A_i)) = MC_i \frac{\eta_i}{\eta_i - 1}, \tag{4''}$$

while the SOC is still (A.1).

Equilibrium allocation. Equation (5) is not valid anymore, but one can obtain an analogous expression using (4”) for any active firms i and j:

$$
\begin{aligned}
\frac{MC_i}{1-\tau(A_i)}(\eta - s_i)^{-1} =& \frac{MC_j}{1-\tau(A_j)}(\eta - s_j)^{-1} \\
s_i - s_j = & \left[1 - \frac{MC_i/(1-\tau(A_i))}{MC_j/(1-\tau(A_j))}\right](\eta - s_j) = \left(1 - \frac{B_j}{B_i}\right)(\eta - s_j), \qquad (5”)
\end{aligned}
$$

where $B_i \equiv A_i\,(1-\tau(A_i))$. In the following, let $\underline{\underline{B}} \equiv \min_i\{B_i\}$ and $\overline{B} \equiv \max_i\{B_i\}$, with $0 < \underline{\underline{B}} < \overline{B} < +\infty$. We again discard any equilibrium in which a non-active firm has a lower marginal cost than an active firm, considering here the adjusted marginal cost $\frac{MC_i}{1-\tau(A_i)} = \left(\frac{r}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\frac{1}{B_i}$. Hence, analogously to the baseline model, in the unique refined equilibrium, there exists a firm with adjusted productivity $\underline{B}$ serving as the cutoff for active firms, such that firm i is active if and only if $B_i \geq \underline{B}$.[41]

In this context, following the same steps of the baseline model but using (5”) instead of (5), one can easily show that

$$
s(B_i) = \frac{1}{N_a \mathrm{E}_a\left(B_i/B\right)} + \eta\left(1 - \frac{1}{\mathrm{E}_a\left(B_i/B\right)}\right), \qquad \text{(A.2”)}
$$

where now $\mathrm{E}_a\left(h(A)\right) \equiv \mathrm{E}\left(h(A)|B \geq \underline{B}\right)$.

We suppose again inefficient technologies are common knowledge, with $\underline{\underline{A}} \to 0$ and $N \to \infty$. Furthermore, although firms may receive subsidies ($\tau_i < 0$), we assume these subsidies do not vanish the dispersion in B in the sense that $(A_i \to 0) \to (B_i \to 0)$. As a result, low-productivity firms will not be active, and thus, the market share $s(\underline{B})$ should be relatively low. Assuming it is approximately zero and using (5”) and (A.2”), one can see that

$$
\begin{aligned}
s(B_i) \approx & \eta\left(1 - \underline{B}/B_i\right) \qquad (6”) \\
\eta \approx & \frac{1}{N_a\left[1 - \mathrm{E}_a\left(\underline{B}/B\right)\right]}. \qquad (7”)
\end{aligned}
$$

Aggregate productivity and misallocation. Equations (8), (A.3), (A.4), (A.5), (9), and (10) remain valid, with $s(A_i)$ replaced by $s(B_i)$. However, allocative efficiency Ω is not given anymore by (11). To obtain the new expression, just plug (6”) and (7”) into (10):

$$
\begin{aligned}
\frac{1}{\overline{A}\Omega} \approx & \frac{\mathrm{E}_a(1/A) - \mathrm{E}_a\left[\underline{B}/(A\times B)\right]}{1 - \mathrm{E}_a\left(\underline{B}/B\right)} = \frac{\mathrm{E}_a(\underline{A}/A) - \mathrm{E}_a\left[(\underline{A}\times\underline{B})/(A\times B)\right]}{\underline{A}\left[1 - \mathrm{E}_a\left(\underline{B}/B\right)\right]} \\
\Omega \approx & \frac{(\underline{A}/\overline{A})\left[1 - \mathrm{E}_a\left(\underline{B}/B\right)\right]}{\mathrm{E}_a(\underline{A}/A) - \mathrm{E}_a\left[(\underline{A}\times\underline{B})/(A\times B)\right]} = \frac{\mathrm{E}_a\left[(\underline{A}/\overline{A})(1 - \underline{B}/B)\right]}{\mathrm{E}_a\left[(\underline{A}/A)(1 - \underline{B}/B)\right]}. \qquad (11”)
\end{aligned}
$$

Average markup. The cost-weighted average markup μ no longer follows Equation (12). How-

[41] Similarly to the baseline discrete model, for the sake of clarity and convenience, we rule out situations where not all firms with productivity $\underline{B}$ are active. As before, this results in just a low loss of generality.

ever, given (i) (9) with $s(A_i)$ replaced by $s(B_i)$, (ii) (4"), (iii) that the active firm with adjusted productivity $\underline{B}$ has markup approximately equal to one, and (iv) the Cobb-Douglas marginal cost function, it is easy to show that

$$\mu = \sum_{i=1}^{N} \theta_i \mu_i \approx \sum_{i=1}^{N} \left[\overline{A}\Omega \frac{s(B_i)}{A_i} \right] \left[\frac{\left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{w}{1-\alpha}\right)^{1-\alpha} \frac{1}{\underline{B}}}{\left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{w}{1-\alpha}\right)^{1-\alpha} \frac{1}{B_i}} \right] = \frac{\overline{A}\Omega}{\underline{B}} \left(1 - \tilde{\tau}\right), \tag{12"}$$

where $\tilde{\tau} \equiv \sum_{i=1}^{N} s(B_i)\tau(A_i)$ is the sales-weighted average tax rate.

Average tax rate. We have already derived all equations analogous to the baseline model.[42] However, this model has an additional equation, for the sales-weighted average tax rate $\tilde{\tau}$:

$$\tilde{\tau} \approx \frac{\mathrm{E}_a \left[\tau(A)(1 - \underline{B}/B)\right]}{\mathrm{E}_a \left(1 - \underline{B}/B\right)}, \tag{F.3}$$

which can be easily obtained by plugging (6") and (7") into $\tilde{\tau} \equiv \sum_{i=1}^{N} s(B_i)\tau(A_i)$.

F.2.2 Remarks on the quantification strategy

We briefly discuss the quantification of this model under two scenarios. First, assuming the function τ is known, one can essentially follow the strategy of Section 3, but using (11") and (12") instead of (11) and (12). Second, suppose the function τ is known except for a parameter, but the average tax rate $\tilde{\tau}$ is observed. In this case, we would need to change the calibration algorithm, since we now should obtain the unknown parameter of the function τ, $\underline{A}$, and $\overline{A}$ by matching Equations (11"), (12"), and also (F.3) to data.

Regarding these empirical strategies, three final comments are due. First, in Appendix D, we discuss necessary and sufficient conditions for the existence of a solution for the calibration algorithm. We do not make an analogous evaluation here, and thus, a solution for the new calibration problems may not exist, at least for some functions τ. Second, $\mu = \frac{1-\alpha}{LS}$ is still valid here, but now $LS \equiv \frac{Hw}{p\sum_{i=1}^{N}(1-\tau_i)Y_i} = \frac{Hw}{Yp(1-\tilde{\tau})}$ is the labor share of national income net of revenue tax. So, in matching (12") to $\frac{1-\alpha}{LS}$, one can ignore the average tax rate $\tilde{\tau}$ and match $\frac{\overline{A}\Omega}{\underline{B}}$ to $\frac{1-\alpha}{LS_g}$, where $LS_g \equiv \frac{Hw}{Yp}$ is the labor share of national gross income. Third, we use here the baseline Cobb-Douglas production function (2). However, in the spirit of Appendix F.1, one can employ an arbitrary well-behaved production function with M factors of production, constant returns to scale, and Hicks-neutral productivity shifter. As before, the only difference would be in the computation of TFP $\overline{A}\Omega$ and average markup μ from data.

[42] As expected, these equations reduce to the baseline equations if $\tau_i = 0$ for every firm i.

G Calibrating the capital share parameter for the US using factor income data

In this section, we calibrate α for the US using real-world data. As shown in Appendix A.4, in our model α equals the cost share of capital, that is, $\alpha = \frac{Kr}{Kr+Hw}$.[43] Thus, to calibrate it, we require factor income data, which are not fully available in National Accounts. As pure profit may not be null in our model, capital expenses cannot be calculated as the residual between national and labor income. In other words, we should distinguish capital expenses from pure profit. Barkai (2020) estimates such factor incomes in the US nonfinancial corporate sector for each year between 1984 and 2014. He takes compensation of employees Hw from the National Income and Productivity Accounts (NIPA), but capital expenses Kr are not obtained as a residual. Instead, they are the product of the nominal value of the physical capital stock and a required rate of return, which approximates the leasing cost of one dollar's worth of capital. This required rate is computed in the spirit of Hall and Jorgenson (1967), depending on the cost of borrowing in financial markets, depreciation rates, expected price inflation of capital, and the tax treatment of both capital and debt. Figure G.1 plots the annual estimates for the cost share of capital $\frac{Kr}{Kr+Hw}$. Since the optimal use of factors is probably a better approximation over the long run, we compute its (i) mean, (ii) median, and (iii) steady-state level from an estimated autoregressive process of order one. We find $\alpha = 0.31$ in all three cases, just slightly lower than the standard calibration.

Using this alternative value for α, we recalculate the average markup $\mu = \frac{1-\alpha}{LS}$. The results are presented in Figure G.2. As expected, the over-time pattern does not change, but these markup estimates are higher than the baseline results of Figure 2, by approximately 0.04. For example, in 2012, $\mu = 1.12$ for $\alpha = 1/3$ and $\mu = 1.16$ for $\alpha = 0.31$. Figure G.3 plots the estimates for allocative efficiency Ω based on $\alpha = 0.31$, which are lower and more volatile compared to the baseline results (Figure 3). These outcomes align with the observations depicted in Figure 1, where a higher markup level points to lower and more volatile Ω.

In any case, at least for growth analysis, the differences are not particularly relevant. In Table G.1, we repeat the growth accounting exercise of Table 2, but considering $\alpha = 0.31$. The results are practically the same. You can uncover slightly more visible quantitative differences in the TFP decomposition for $k = 5$ (e.g., for 1995–2000 and 2000–2007). However, even in these cases, the fundamental qualitative assessments remain unchanged.

[43] Setting α as the cost share of capital is conceptually equivalent to calibrating the average markup μ to match the pure profit share of national income, defined as $PS \equiv 1 - \frac{Kr+Hw}{Yp}$. Formally, if $\alpha = \frac{Kr}{Kr+Hw}$, $\mu = \frac{1-\alpha}{LS} = \frac{\frac{Hw}{Kr+Hw}}{\frac{Hw}{Yp}} = \frac{1}{\frac{Kr+Hw}{Yp}} = \frac{1}{1-PS} \leftrightarrow PS = \frac{\mu-1}{\mu}$.

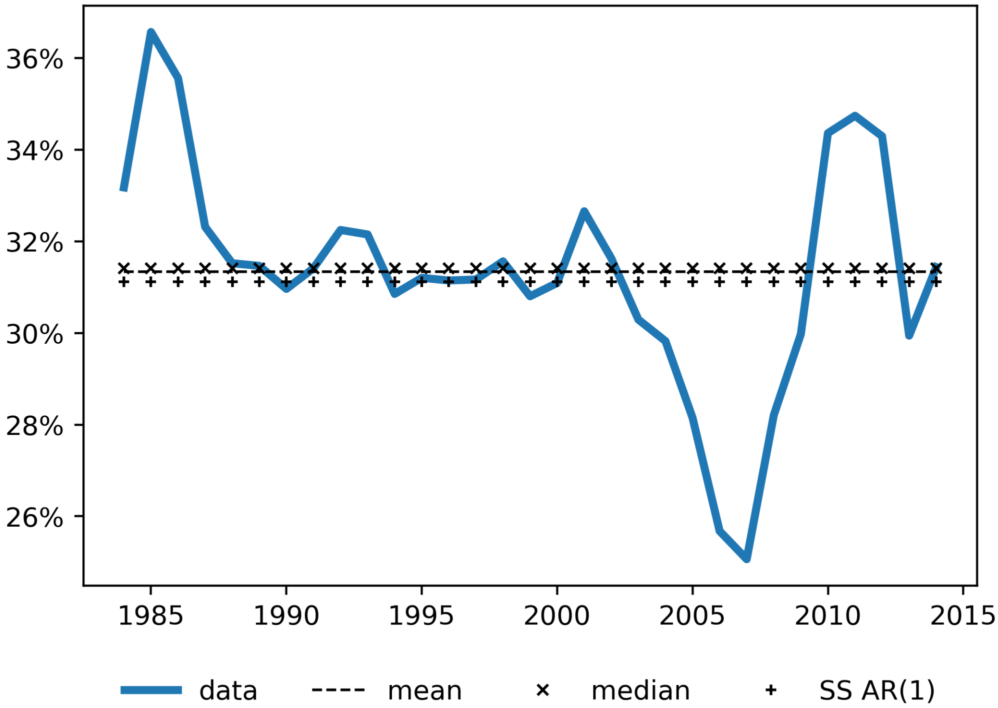

Figure G.1: Cost share of physical capital in the US nonfinancial corporate sector. *Source*: computations based on data from Barkai (2020).

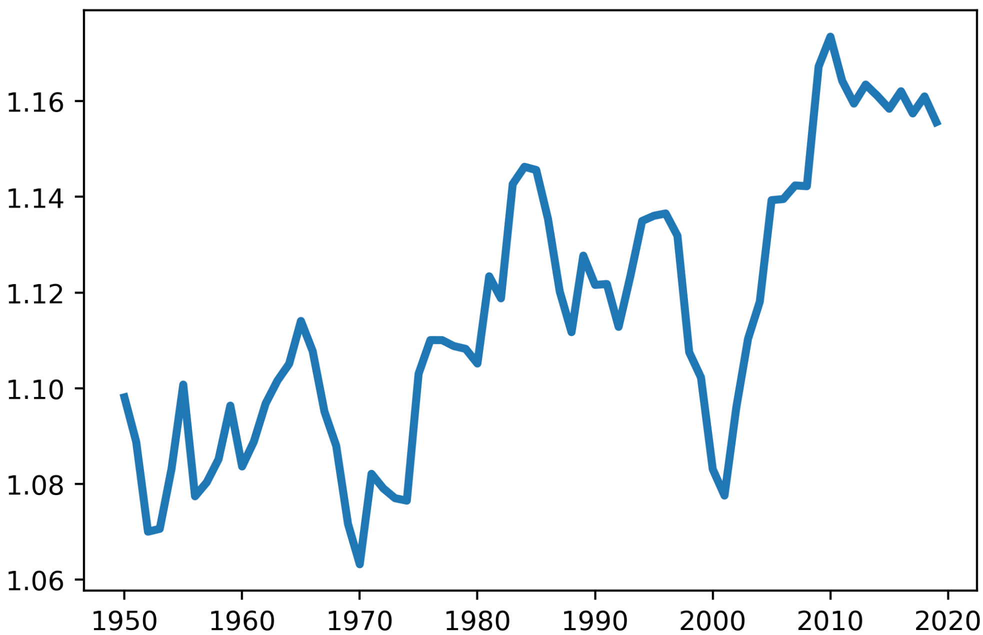

Figure G.2: Average markup in the United States for $\alpha = 0.31$.

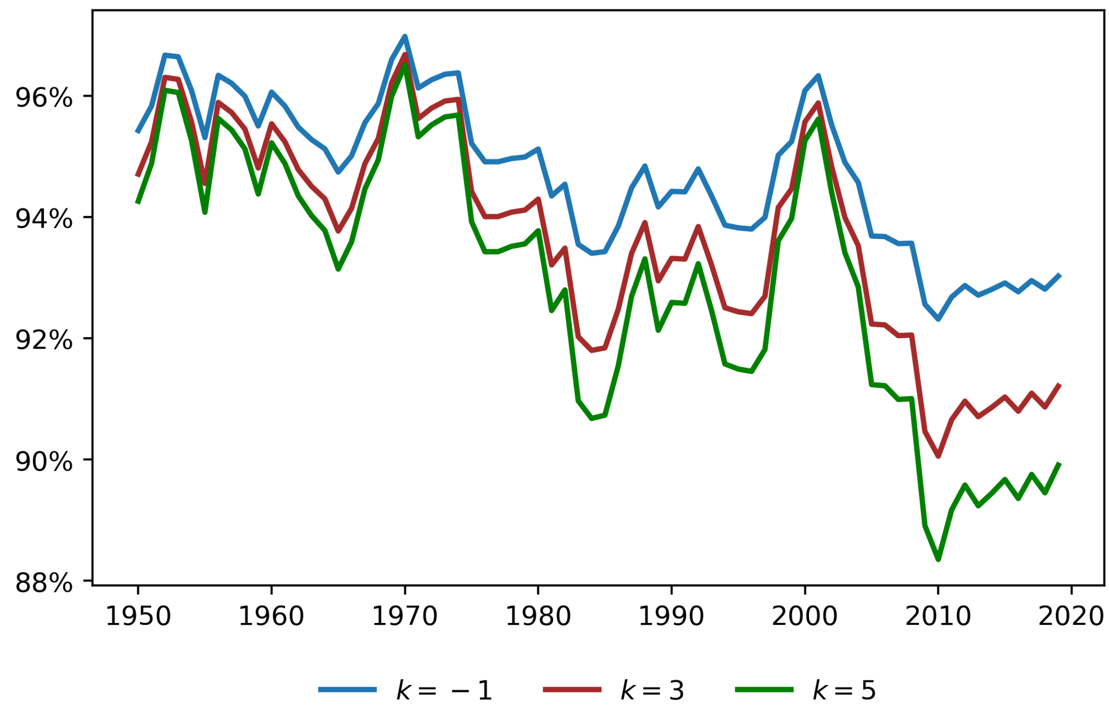

Figure G.3: Allocative efficiency in the United States for $\alpha = 0.31$.

Table G.1: Growth accounting for the United States using $\alpha = 0.31$

Period	$\frac{Y}{L}$	Y/L components			Labor-aug. TFP $\left(\overline{A}\Omega\right)^{\beta}$ components					
					$k=-1$		$k=3$		$k=5$	
		$\left(\frac{K}{Y}\right)^{\alpha\beta}$	$\frac{H}{L}$	$\left(\overline{A}\Omega\right)^{\beta}$	$\overline{A}^{\beta}$	Ω^{β}	$\overline{A}^{\beta}$	Ω^{β}	$\overline{A}^{\beta}$	Ω^{β}
1954–2019	1.9	0.2	0.5	1.2	1.2	-0.1	1.3	-0.1	1.3	-0.1
1954–2013	1.9	0.2	0.6	1.2	1.3	-0.1	1.3	-0.1	1.3	-0.2
1954–1973	2.6	-0.0	1.0	1.7	1.7	0.0	1.7	0.0	1.7	0.0
1973–1990	1.3	0.3	0.5	0.5	0.7	-0.2	0.8	-0.2	0.8	-0.3
1990–1995	1.6	0.2	0.5	0.9	1.1	-0.2	1.2	-0.3	1.3	-0.3
1995–2000	2.2	0.2	0.3	1.7	1.0	0.7	0.8	1.0	0.6	1.2
2000–2007	2.2	0.3	0.3	1.5	2.1	-0.6	2.3	-0.8	2.5	-0.9
2007–2013	1.3	0.4	0.3	0.6	0.8	-0.2	0.9	-0.4	1.0	-0.5
2013–2019	1.0	-0.1	0.1	1.0	0.9	0.1	0.8	0.1	0.8	0.2

Note: Logarithmic approximation of average annual growth rates (in percent).

H Exploring the data sets for level analyses

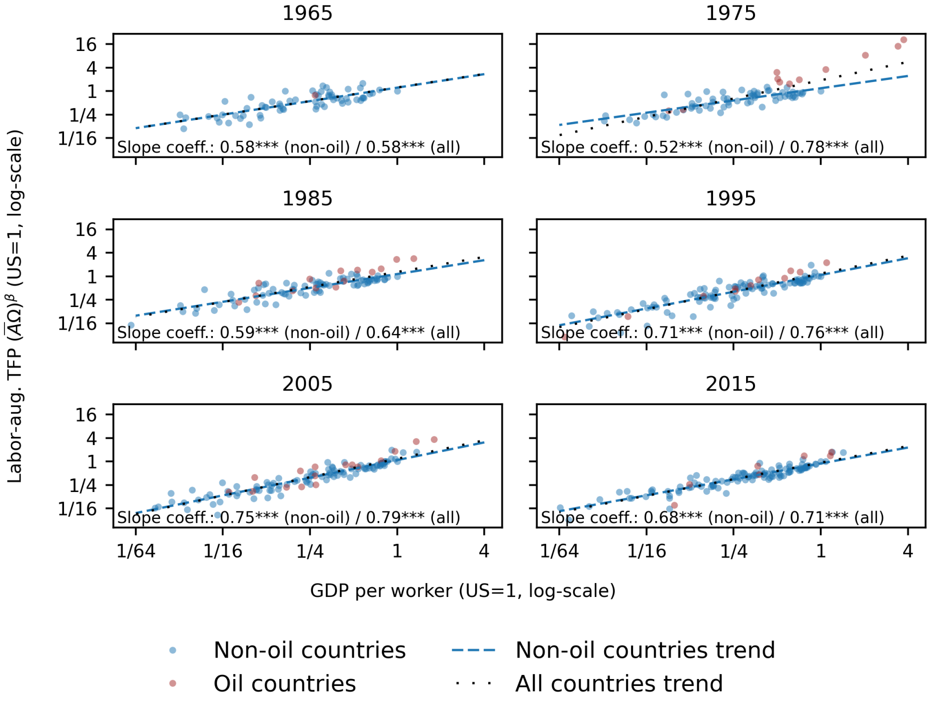

Figure H.1: Labor-aug. TFP vs. GDP per worker – level data set with greatest coverage and $k = -1$.

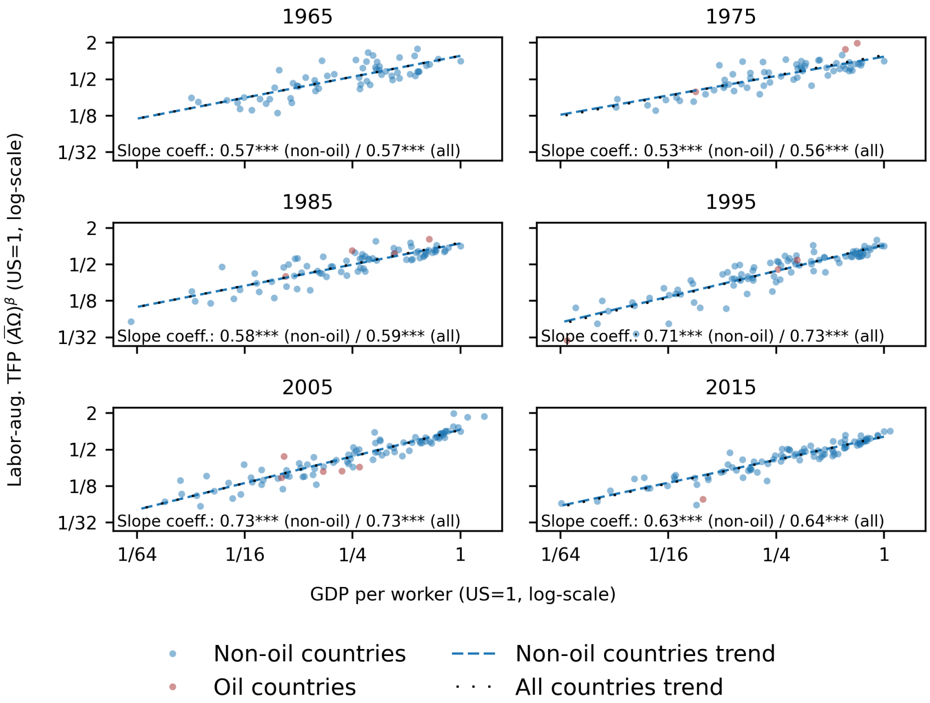

Figure H.2: Labor-aug. TFP vs. GDP per worker – level data set with greatest coverage and $k = 3$.

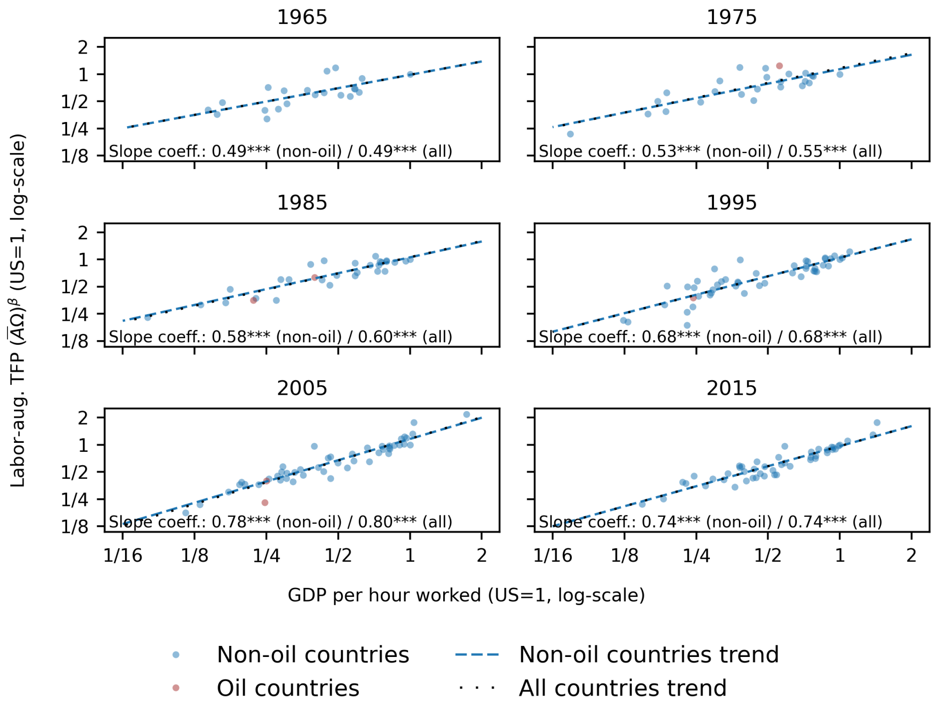

Figure H.3: Labor-aug. TFP vs. GDP per hour worked – best proxies for level analyses and $k = -1$.

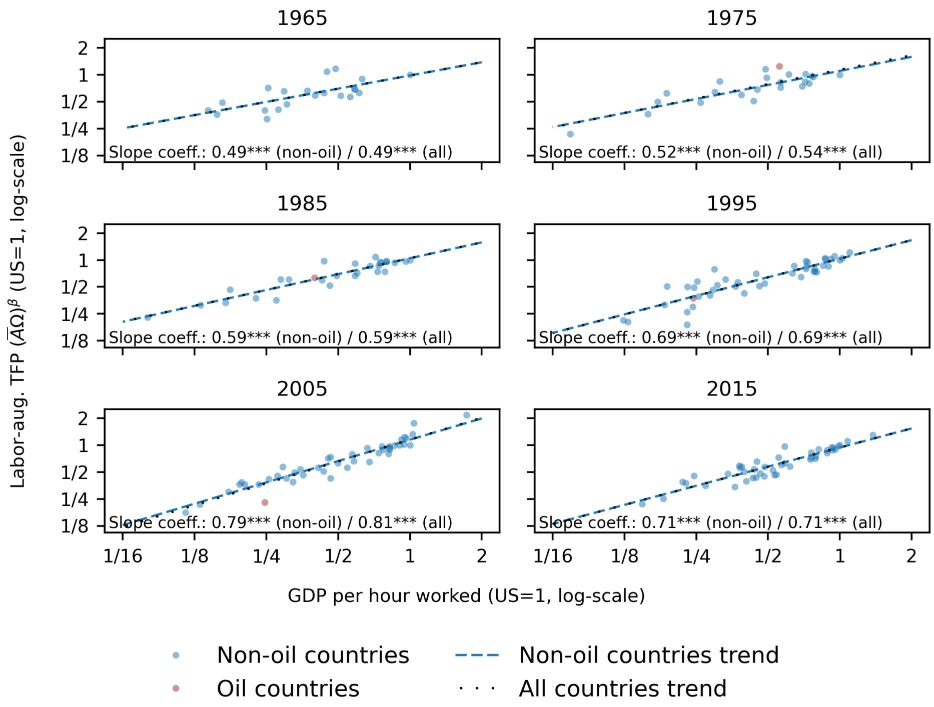

Figure H.4: Labor-aug. TFP vs. GDP per hour worked – best proxies for level analyses and $k = 3$.

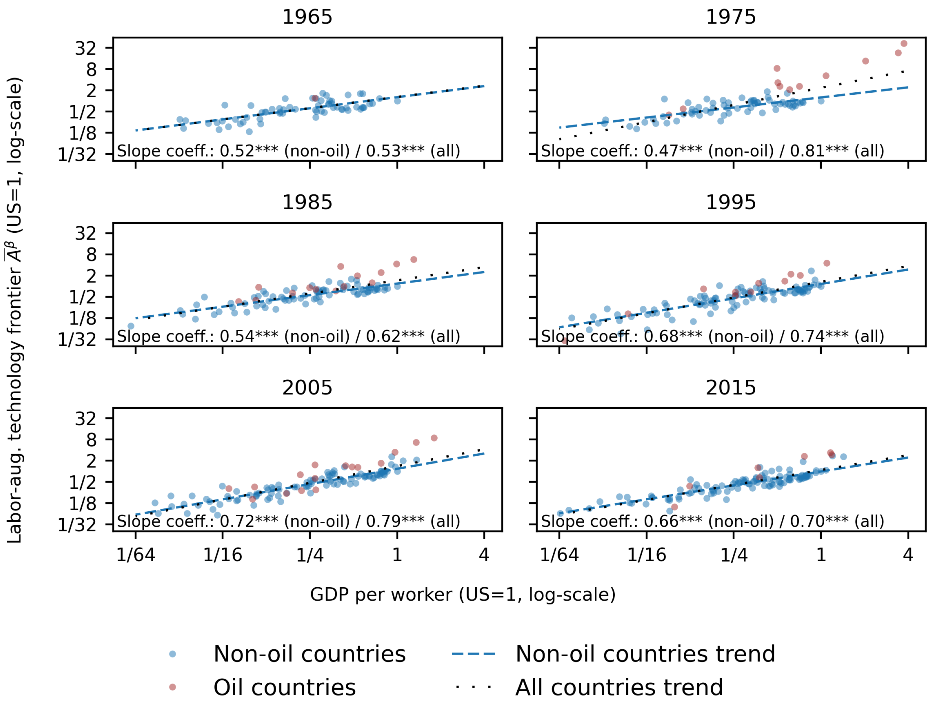

Figure H.5: Labor-aug. technology frontier vs. GDP per worker – level data set with greatest coverage and $k = -1$.

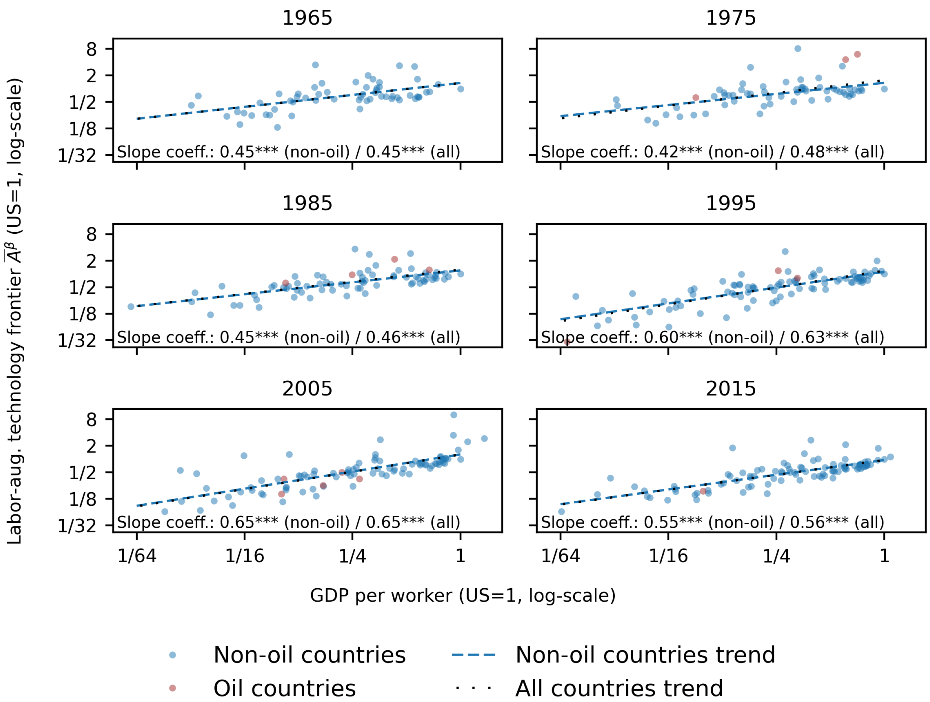

Figure H.6: Labor-aug. technology frontier vs. GDP per worker – level data set with greatest coverage and $k = 3$.

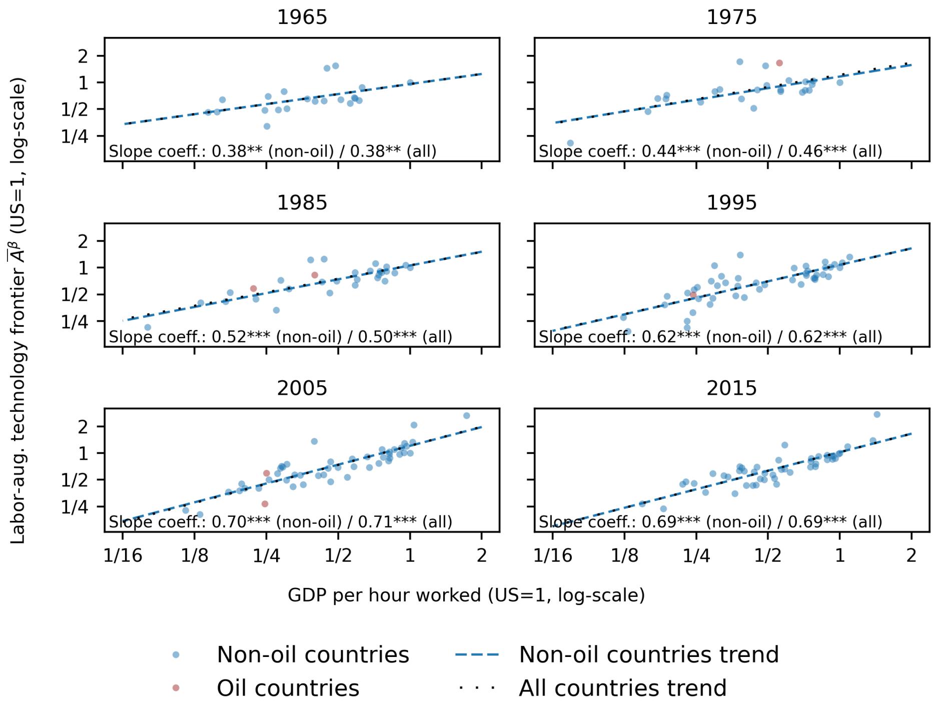

Figure H.7: Labor-aug. technology frontier vs. GDP per hour worked – best proxies for level analyses and $k = -1$.

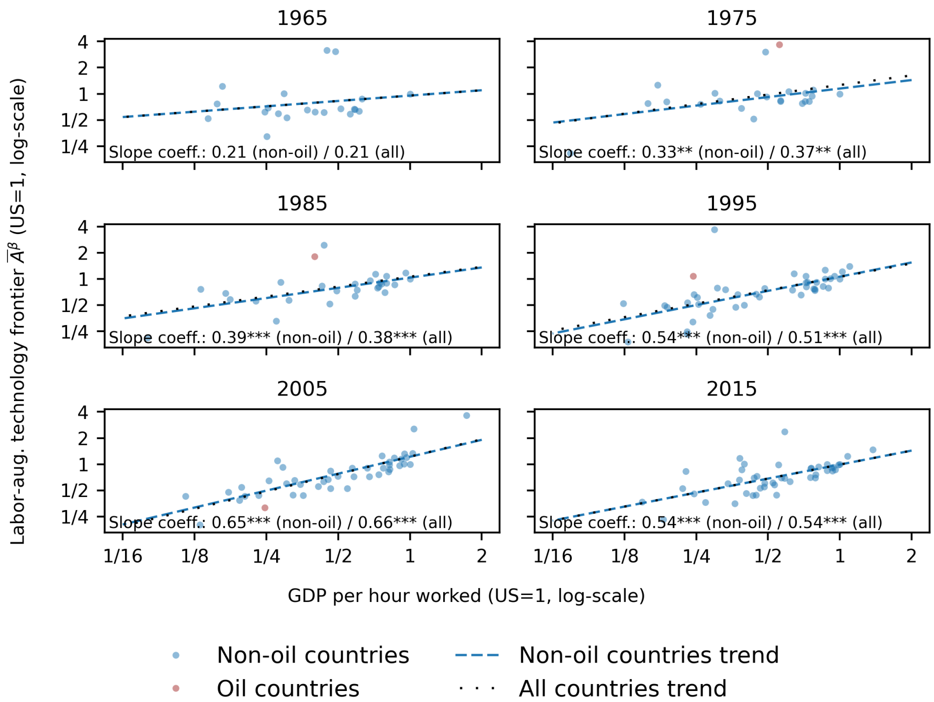

Figure H.8: Labor-aug. technology frontier vs. GDP per hour worked – best proxies for level analyses and $k = 3$.

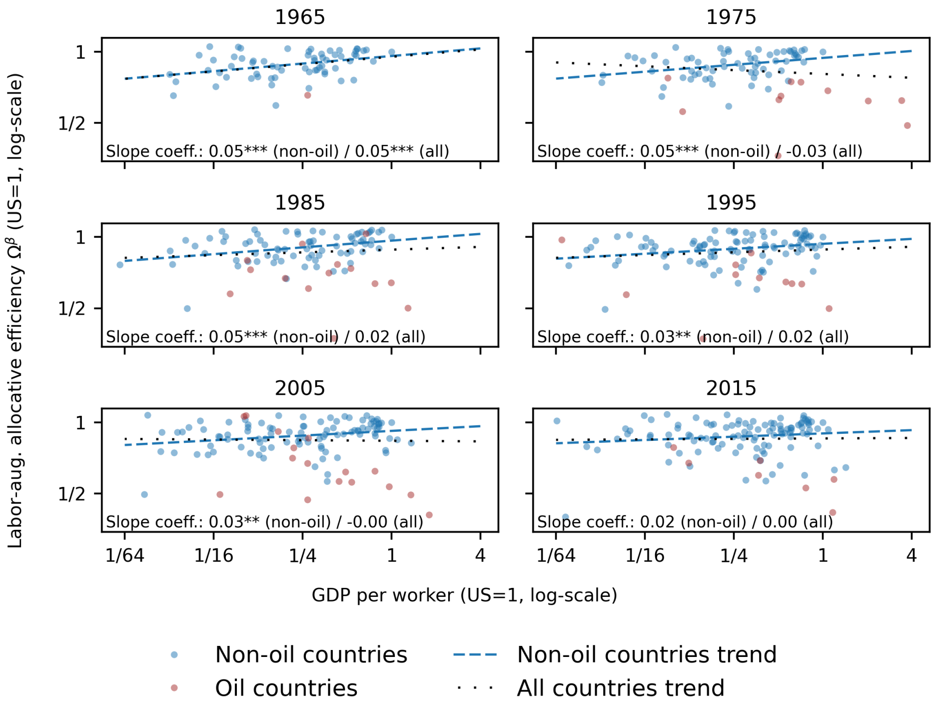

Figure H.9: Labor-aug. allocative efficiency vs. GDP per worker – level data set with greatest coverage and $k = -1$.

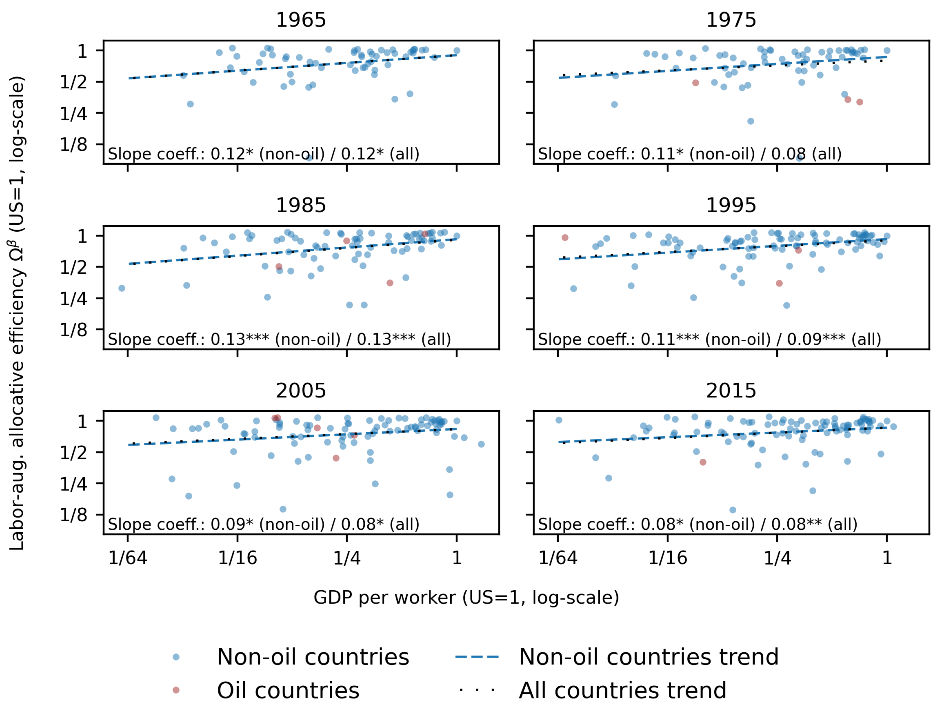

Figure H.10: Labor-aug. allocative efficiency vs. GDP per worker – level data set with greatest coverage and $k = 3$.

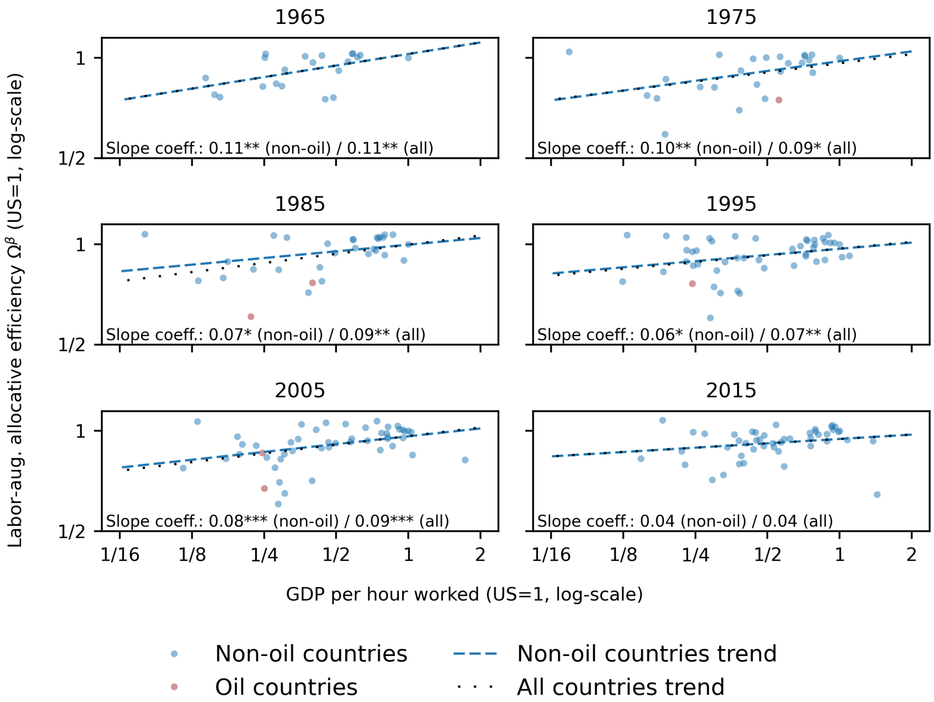

Figure H.11: Labor-aug. allocative efficiency vs. GDP per hour worked – best proxies for level analyses and $k = -1$.

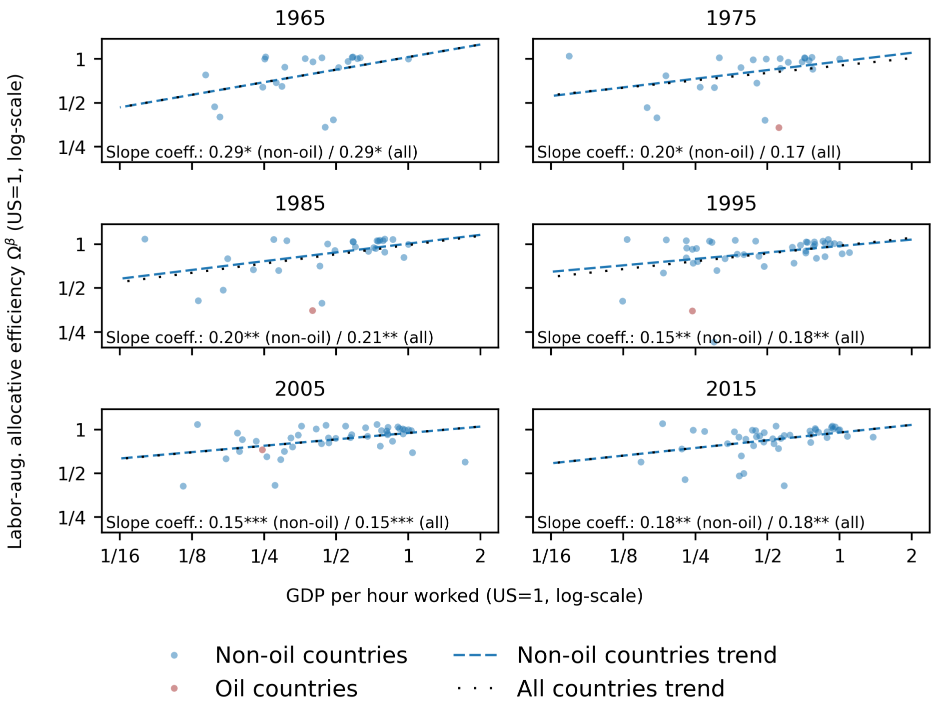

Figure H.12: Labor-aug. allocative efficiency vs. GDP per hour worked – best proxies for level analyses and $k = 3$.

I Exploring the data sets for growth analyses

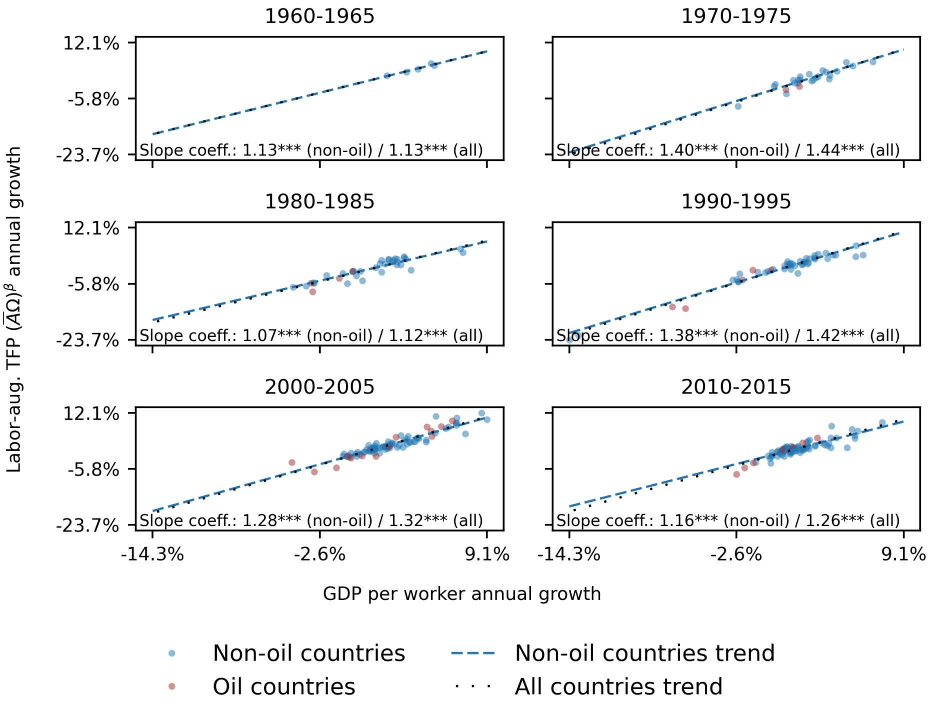

Figure I.1: Labor-aug. TFP growth vs. GDP per worker growth – growth data set with greatest coverage and $k = -1$.

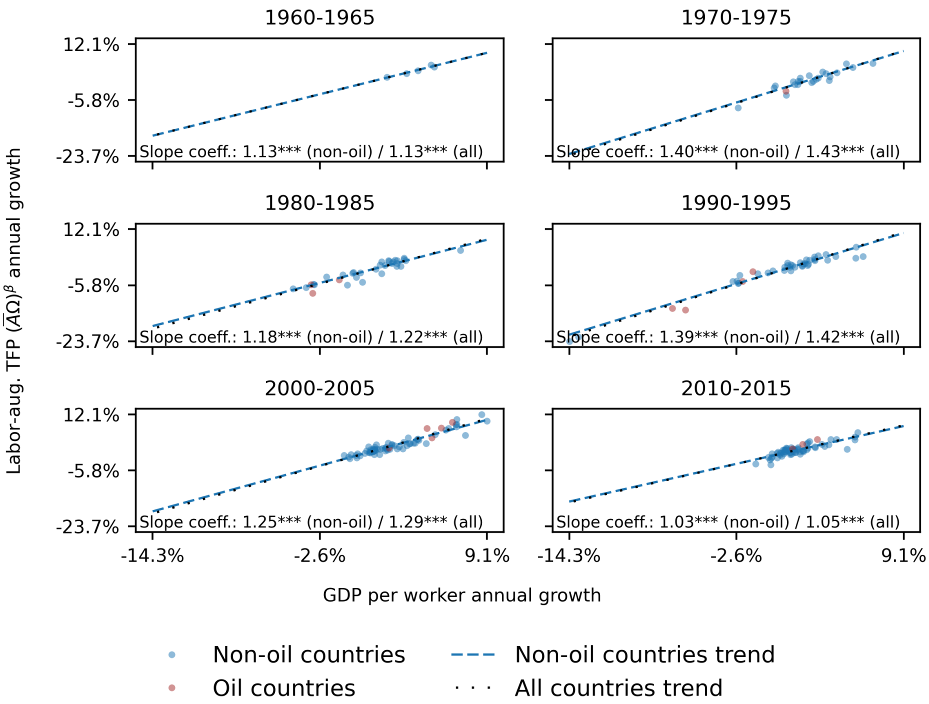

Figure I.2: Labor-aug. TFP growth vs. GDP per worker growth – growth data set with greatest coverage and $k = 3$.

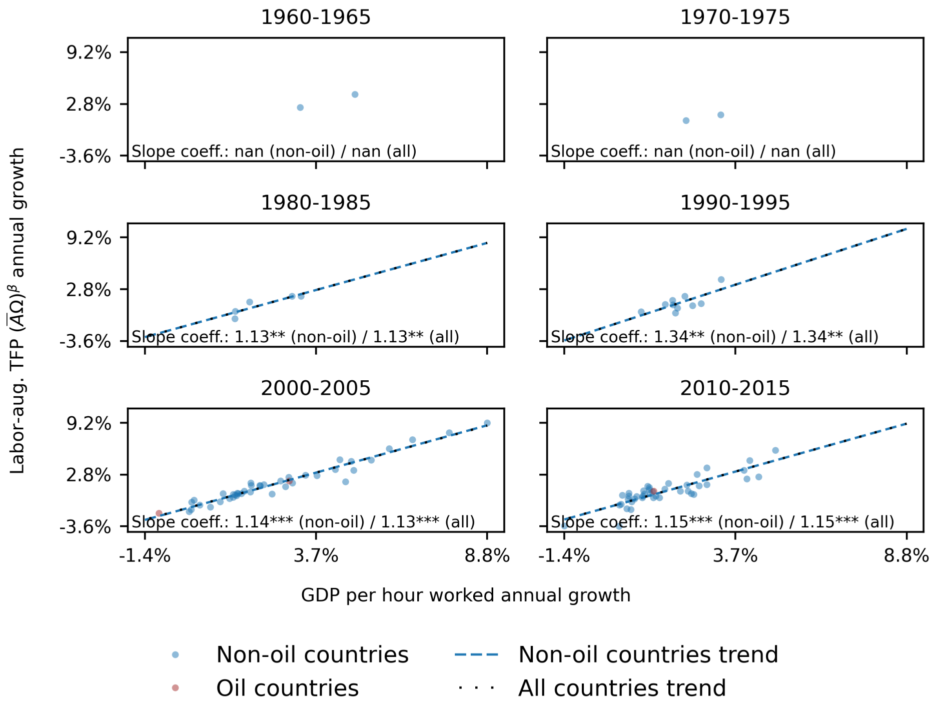

Figure I.3: Labor-aug. TFP growth vs. GDP per hour worked growth – best proxies for growth analyses and $k = -1$.

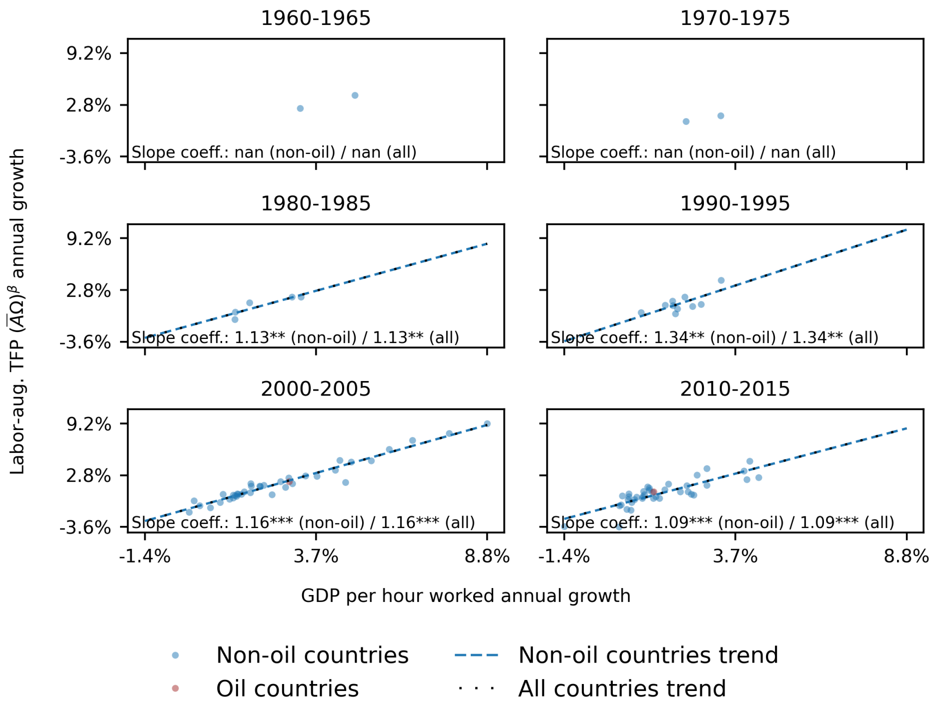

Figure I.4: Labor-aug. TFP growth vs. GDP per hour worked growth – best proxies for growth analyses and $k = 3$.

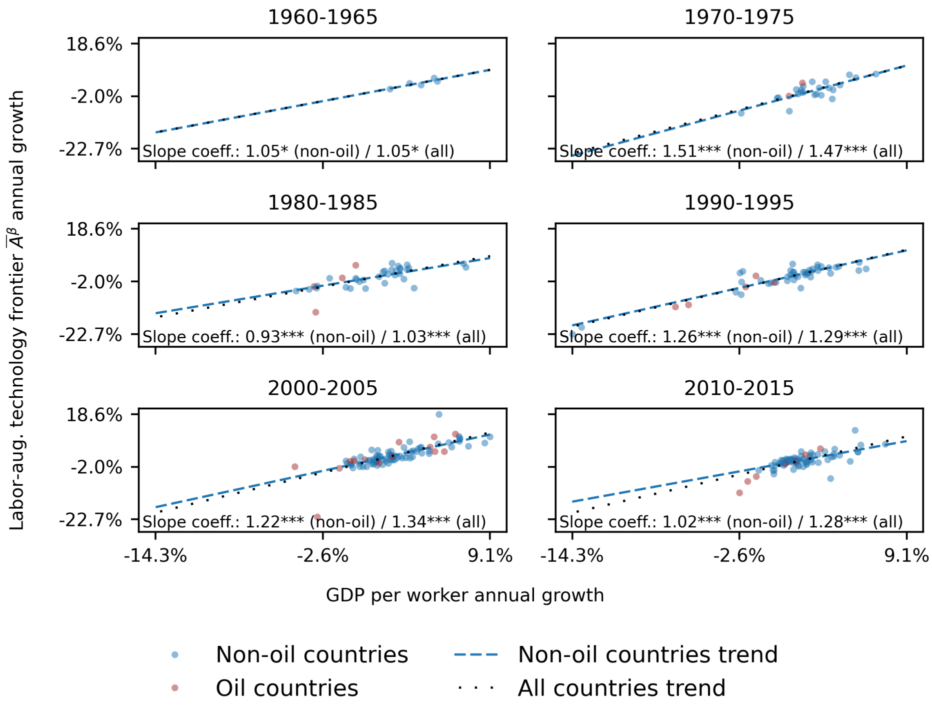

Figure I.5: Labor-aug. technology frontier growth vs. GDP per worker growth – growth data set with greatest coverage and $k = -1$.

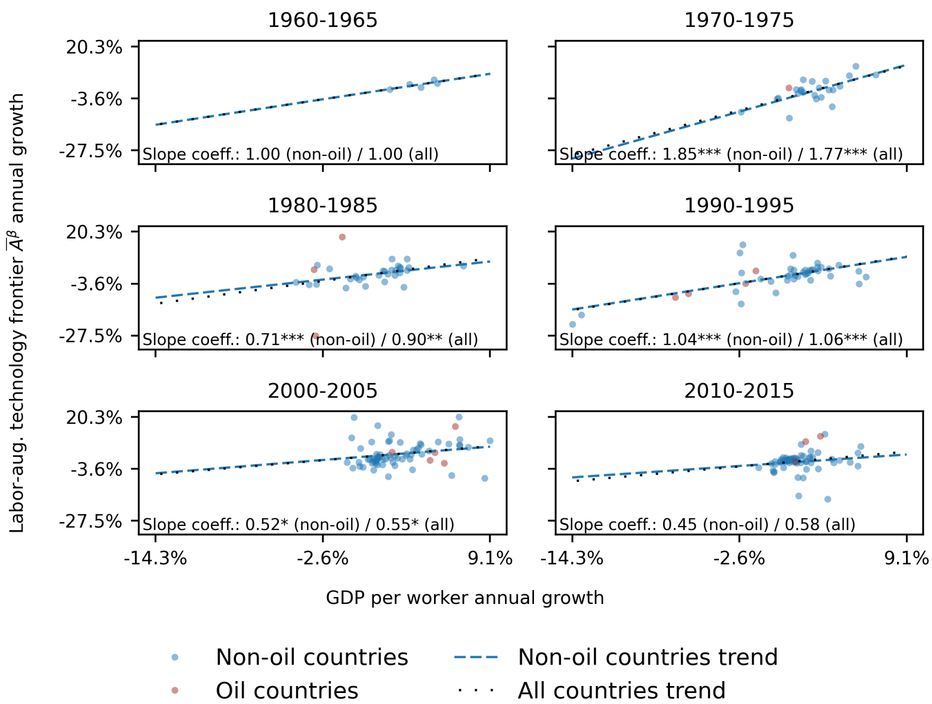

Figure I.6: Labor-aug. technology frontier growth vs. GDP per worker growth – growth data set with greatest coverage and $k = 3$.

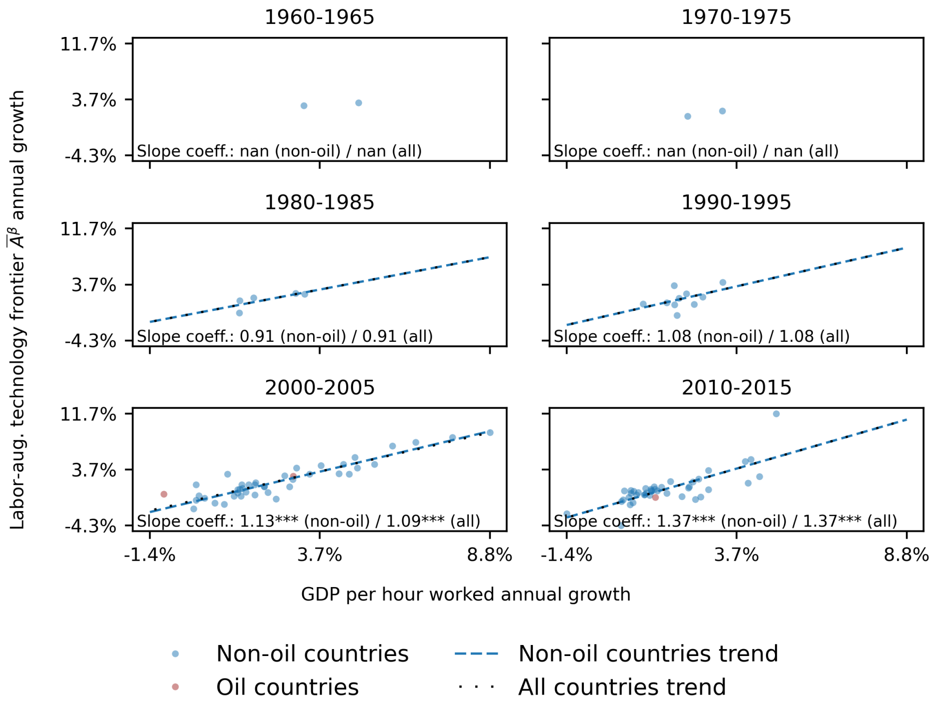

Figure I.7: Labor-aug. technology frontier growth vs. GDP per hour worked growth – best proxies for growth analyses and $k = -1$.

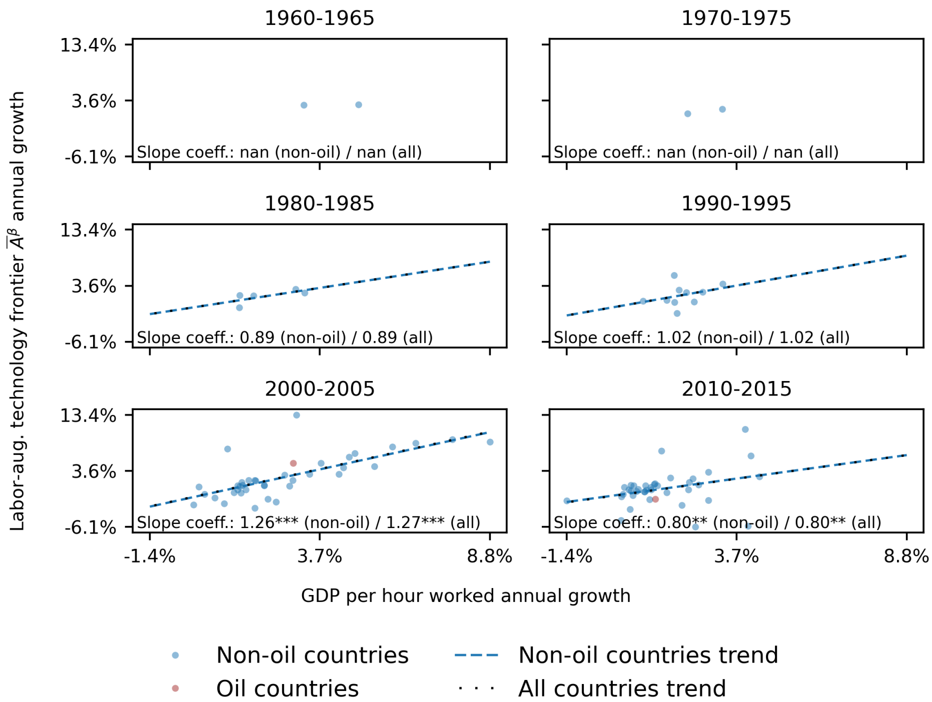

Figure I.8: Labor-aug. technology frontier growth vs. GDP per hour worked growth – best proxies for growth analyses and $k = 3$.

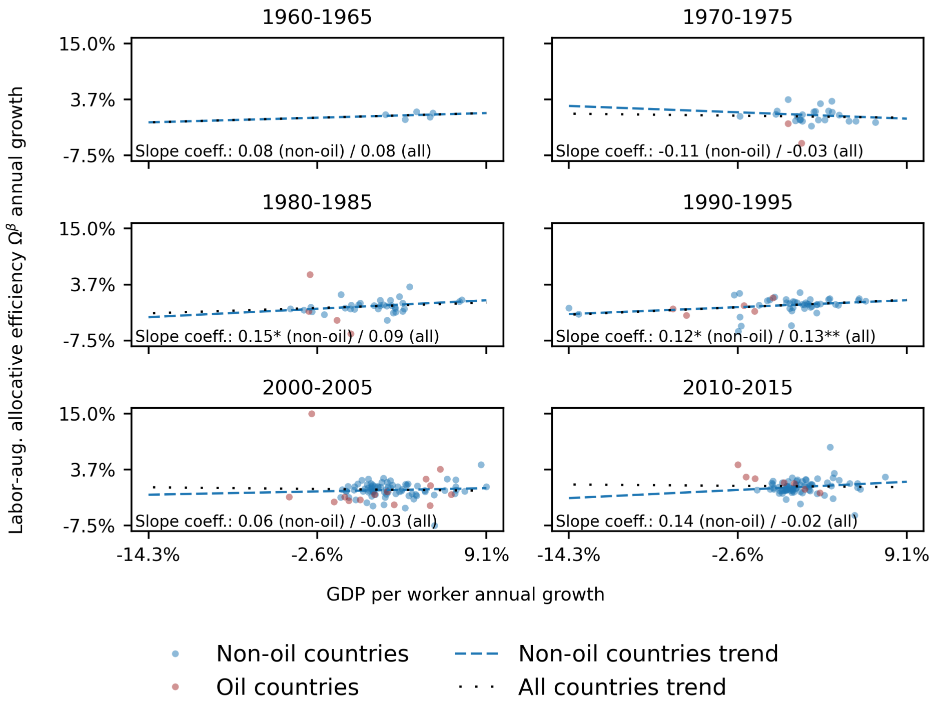

Figure I.9: Labor-aug. allocative efficiency growth vs. GDP per worker growth – growth data set with greatest coverage and $k = -1$.

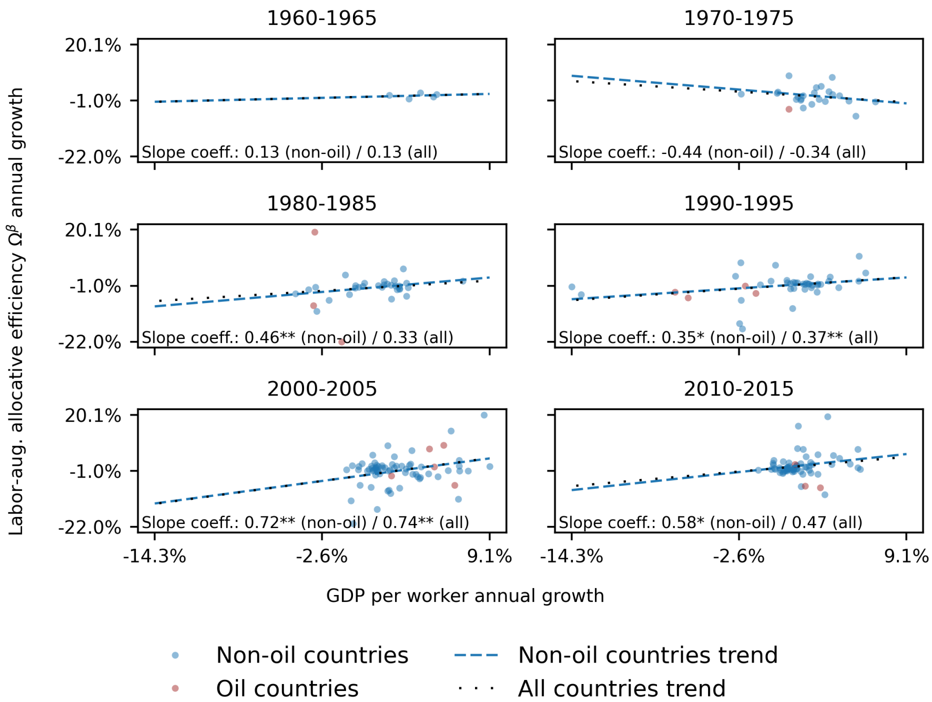

Figure I.10: Labor-aug. allocative efficiency growth vs. GDP per worker growth – growth data set with greatest coverage and $k = 3$.

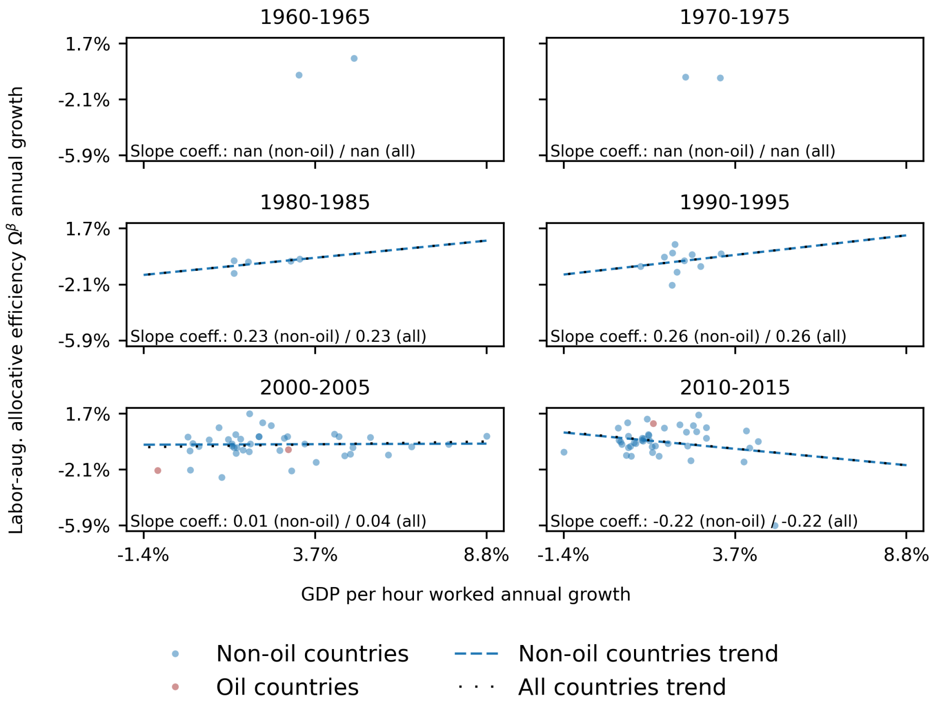

Figure I.11: Labor-aug. allocative efficiency growth vs. GDP per hour worked growth – best proxies for growth analyses and $k = -1$.

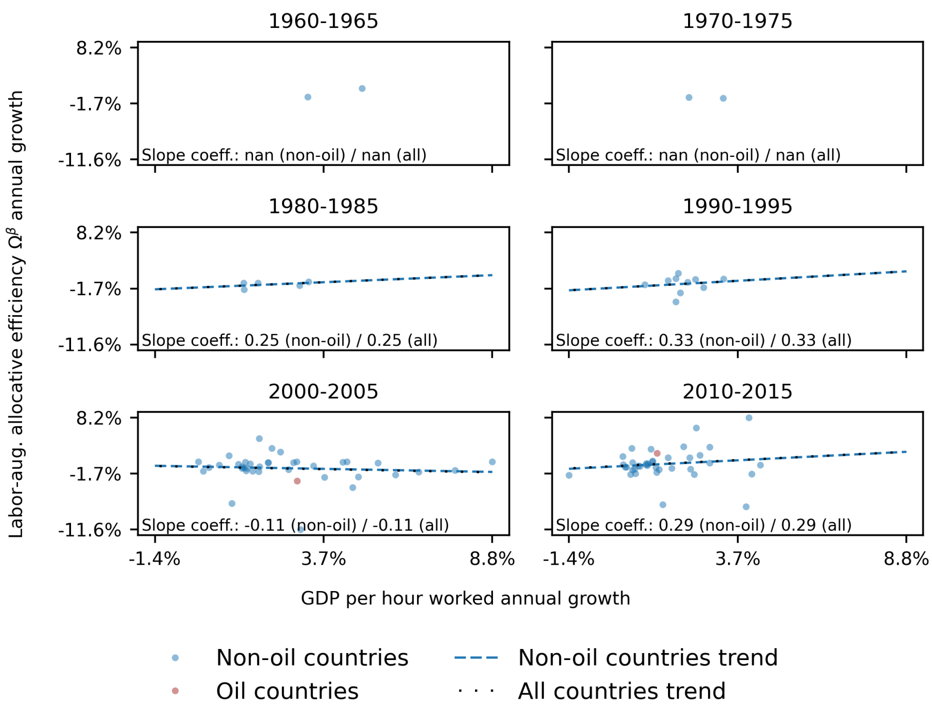

Figure I.12: Labor-aug. allocative efficiency growth vs. GDP per hour worked growth – best proxies for growth analyses and $k = 3$.

J Measures of success for different samples

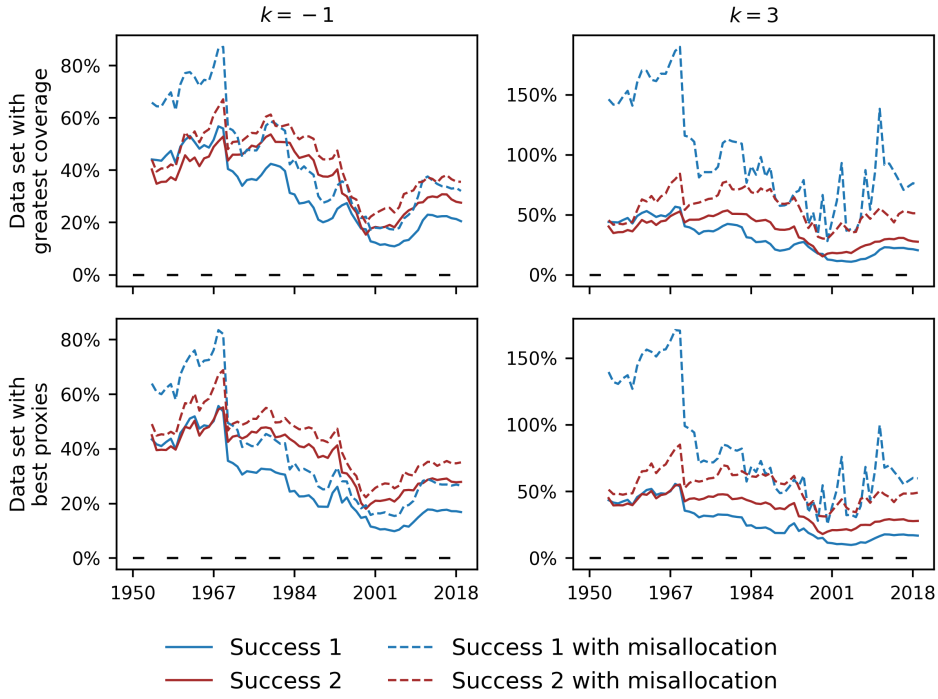

Figure J.1: Measures of success, same non-oil countries' sample across plots in each given year.

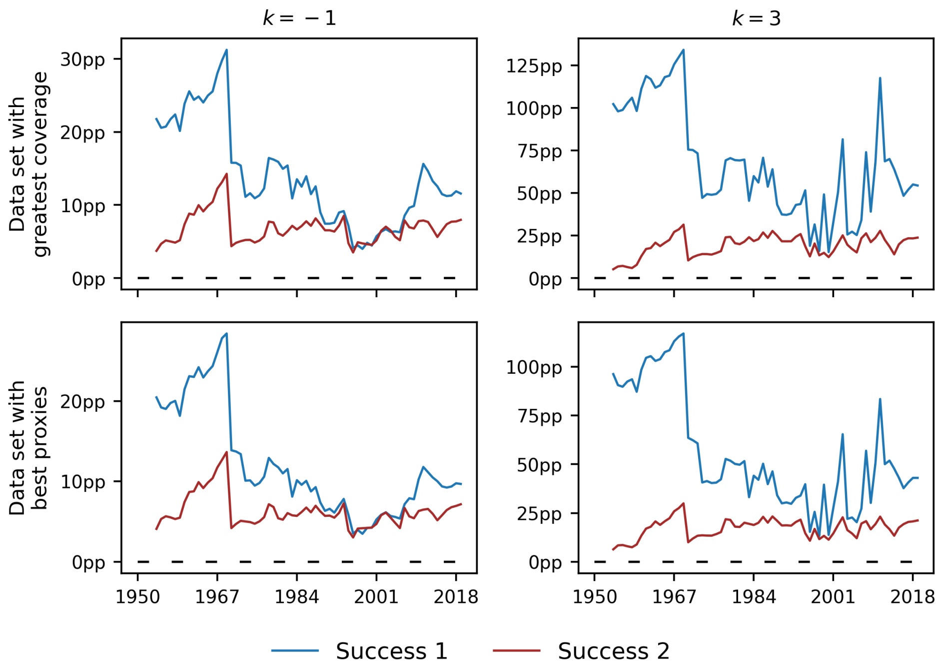

Figure J.2: Success gains due to misallocation, same non-oil countries' sample across plots in each given year.

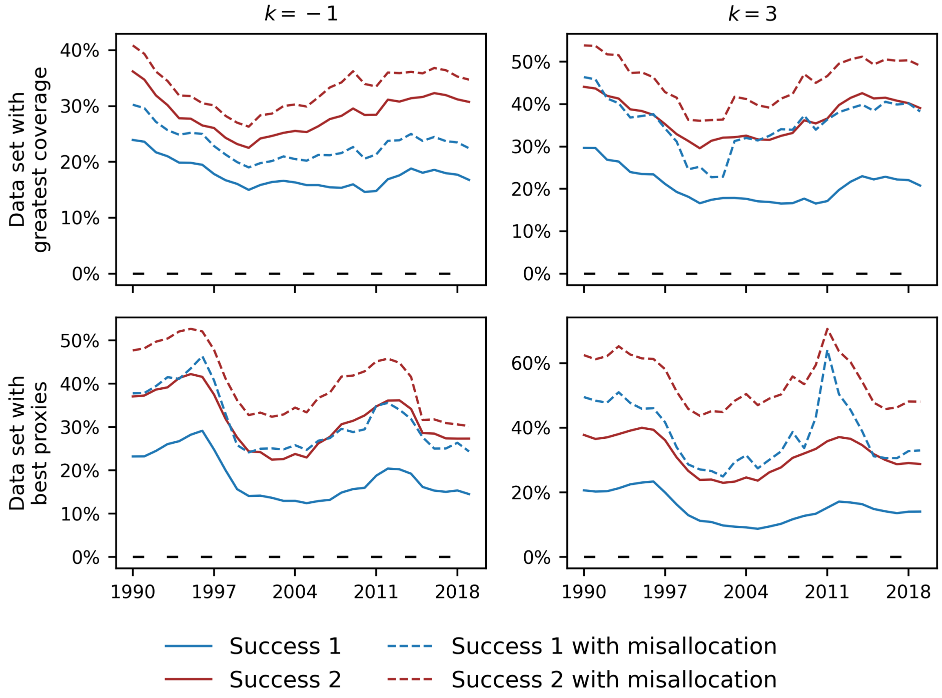

Figure J.3: Measures of success, time-invariant non-oil countries' sample.

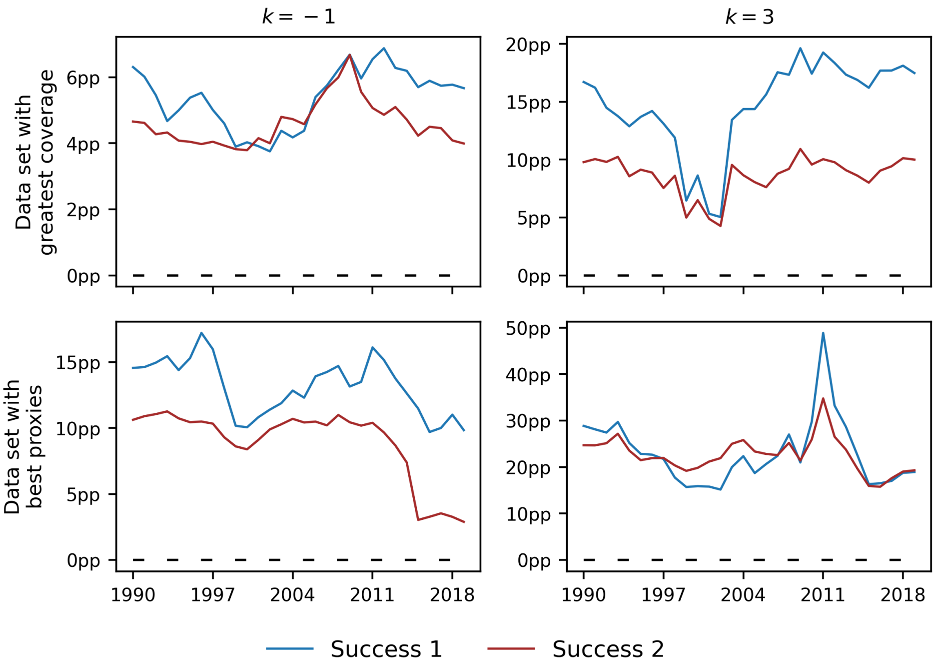

Figure J.4: Success gains due to misallocation, time-invariant non-oil countries' sample.

References

Aghion, P. and P. Howitt (1992). A model of growth through creative destruction. *Econometrica 60*(2), 323–351.

Atkeson, A. and A. Burstein (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review 98*(5), 1998–2031.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics 135*(2), 645–709.

Baqaee, D. R. and E. Farhi (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics 135*(1), 105–163.

Barkai, S. (2020). Declining labor and capital shares. *The Journal of Finance 75*(5), 2421–2463.

Basu, S. and J. G. Fernald (2002). Aggregate productivity and aggregate technology. *European Economic Review 46*(6), 963–991.

Bergeaud, A., G. Cette, and R. Lecat (2018). The role of production factor quality and technology diffusion in twentieth-century productivity growth. *Cliometrica 12*(1), 61–97.

Berger, D., K. Herkenhoff, and S. Mongey (2022). Labor market power. *American Economic Review 112*(4), 1147–93.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum (2003). Plants and productivity in international trade. *American Economic Review 93*(4), 1268–1290.

Byrne, D. M., S. D. Oliner, and D. E. Sichel (2013, Spring). Is the information technology revolution over? *International Productivity Monitor 25*, 20–36.

Calligaris, S., C. Criscuolo, and L. Marcolin (2018). Mark-ups in the digital era. *OECD Science, Technology and Industry Working Papers* (2018/10).

Caselli, F. (2005). Accounting for cross-country income differences. In P. Aghion and S. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 1A, Chapter 9, pp. 679–741. Elsevier.

Caselli, F. (2016). Accounting for cross-country income differences: Ten years later. *World Bank Background Paper*.

Crafts, N. and P. Woltjer (2021). Growth accounting in economic history: findings, lessons and new directions. *Journal of Economic Surveys 35*(3), 670–696.

De Loecker, J. and J. Eeckhout (2018). Global market power. *NBER Working Paper Series* (24768).

De Loecker, J., J. Eeckhout, and S. Mongey (2021). Quantifying market power and business dynamism in the macroeconomy. *NBER Working Paper Series* (28761).

De Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics 135*(2), 561–644.

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review 102*(6), 2437–2471.

Edmond, C., V. Midrigan, and D. Y. Xu (2015). Competition, markups, and the gains from international trade. *American Economic Review 105*(10), 3183–3221.

Edmond, C., V. Midrigan, and D. Y. Xu (2022). How costly are markups? *Journal of Political Economy*.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The next generation of the Penn World Table. *American Economic Review 105*(10), 3150–3182.

Fernald, J. G. (2015). Productivity and potential output before, during, and after the great recession. *NBER Macroeconomics Annual 29*(1), 1–51.

Gollin, D. (2002). Getting income shares right. *Journal of Political Economy 110*(2), 458–474.

Gordon, R. J. (2000). Does the "new economy" measure up to the great inventions of the past? *Journal of Economic Perspectives 14*(4), 49–74.

Hall, R. E. (1988). The relation between price and marginal cost in US industry. *Journal of Political Economy 96*(5), 921–947.

Hall, R. E. and D. W. Jorgenson (1967). Tax policy and investment behavior. *The American Economic Review 57*(3), 391–414.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing TFP in China and India. *The Quarterly journal of economics 124*(4), 1403–1448.

Hulten, C. R. (2010). Growth accounting. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of the Economics of Innovation*, Volume 2, pp. 987–1031. Elsevier.

Jones, C. I. (2016). The facts of economic growth. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Volume 2, Chapter 1, pp. 3–69. Elsevier.

Klenow, P. J. and A. Rodríguez-Clare (1997). The neoclassical revival in growth economics: Has it gone too far? *NBER macroeconomics annual 12*, 73–114.

Martinez, T. R. and T. T. O. Santos (2024). Disentangling Brazilian TFP: the role of misallocation in recent economic cycles. *Central Bank of Brazil Working Paper Series*. Forthcoming.

Melitz, M. J. and S. J. Redding (2015). Heterogeneous firms and trade. *Handbook of International Economics 4*, 1–54.

Neary, J. P. (2010). Two and a half theories of trade. *The World Economy 33*(1), 1–19.

Peters, M. (2020). Heterogeneous markups, growth, and endogenous misallocation. *Econometrica 88*(5), 2037–2073.

Petrin, A. and J. Levinsohn (2012). Measuring aggregate productivity growth using plant-level data. *The Rand Journal of Economics 43*(4), 705–725.

Prescott, E. C. (1998). Lawrence R. Klein lecture 1997: Needed: A theory of total factor productivity. *International Economic Review 39*(3), 525–551.

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics 11*(4), 707–720.

Restuccia, D. and R. Rogerson (2013). Misallocation and productivity. *Review of Economic Dynamics 16*(1), 1–10.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy 98*(5, Part 2), S71–S102.

Traina, J. (2018). Is aggregate market power increasing? production trends using financial statement. *Stigler Center New Working Paper Series* (17).

Wang, O. and I. Werning (2022). Dynamic oligopoly and price stickiness. *American Economic Review 112*(8), 2815–49.