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

Assessing the role of misallocation in shaping total factor productivity (TFP) has been constrained by the availability of firm-level data. This paper addresses this limitation by developing a tractable Cournot model that uses standard macroeconomic data to quantify market-power-driven misallocation. We implement this framework to decompose aggregate TFP into technology and allocative efficiency components for up to one hundred countries over 1950–2019. This decomposition is employed to revisit key facts of economic growth. First, at the world income frontier, proxied by the United States, misallocation plays a minor role in long-run TFP growth but has sizable effects over shorter horizons. Second, in cross-country development accounting, incorporating misallocation improves our understanding of cross-country income differences, though most income variation remains unexplained. Finally, we find little evidence of convergence in income or in aggregate TFP, with the latter reflecting the lack of convergence in its technology component, as our results do provide some support for convergence in allocative efficiency.

JEL codes: L13, D24, O40, O47.
Keywords: Total factor productivity, misallocation, market power, Cournot competition, development accounting, growth accounting.

This paper should not be reported as representing the views of the Banco Central do Brasil. The views expressed in the papers are those of the authors and do not necessarily reflect those of the Banco Central do Brasil.

*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, both across countries and over time (Klenow and Rodríguez-Clare, 1997; Caselli, 2005, 2016; Jones, 2016; Bergeaud et al., 2018; Crafts and Woltjer, 2021). Yet TFP is typically measured as a residual—famously described as "a measure of our ignorance"—capturing the portion of output not explained by observable inputs. This view has motivated a large literature aiming to uncover the underlying sources of productivity differences, as illustrated by the title of Prescott's (1998) influential paper, "Needed: A Theory of TFP." More recently, Jones (2016) suggests that the misallocation literature has emerged precisely to provide the type of theory Prescott called for. Despite its conceptual appeal, measuring misallocation typically requires firm-level data, which are available only for a limited set of countries, years, and sectors—most often manufacturing. This constraint has hindered more comprehensive assessments of the role of misallocation in shaping TFP.

Against this backdrop, this paper makes two contributions. Methodologically, we develop a tractable Cournot model with heterogeneous firms that quantifies market-power-driven misallocation using standard macroeconomic data. The novel contribution of our framework lies precisely in this reliance on macro aggregates for calibration, which enables a comprehensive assessment of misallocation across both time and space. Empirically, we use the model to decompose TFP into technology and allocative efficiency components for up to one hundred countries over the period 1950–2019. These estimates are incorporated into growth and development accounting exercises to revisit key facts of economic growth.

Our model builds on the oligopoly framework of Edmond et al. (2023), but focuses on a tractable special case in which (i) goods are homogeneous within sectors but highly differentiated across sectors; and (ii) barriers to entry are low. Despite these restrictions, the model remains rich, featuring endogenous entry, firm-level decisions about whether to operate, and oligopolistic competition among heterogeneous firms. Misallocation arises endogenously because marginal products are not equalized across firms: more productive firms command larger market shares, face lower demand elasticities, and therefore charge higher markups. In return, these stronger assumptions yield a parsimonious and tractable framework.

First, all sectors are identical in equilibrium, so the model can be solved for a single representative sector. In contrast, in Edmond et al. (2023), sectors differ ex post due to sampling variation: although firms draw productivity from a common distribution, the small number of firms within each sector generates substantial dispersion in realized sectoral productivity. In our framework, this sampling noise vanishes because equilibrium features a large number of active firms in each sector. Low entry barriers generate substantial firm entry, while the low elasticity of substitution across sectoral goods implies weak competitive pressure from other sectors, allowing many entrants to remain active in equilibrium. Remarkably, however, the large number of firms does not necessarily eliminate strategic interaction. Perfect competition

arises only when the number of entrants becomes large *relative* to the inverse of the elasticity of substitution.

Second, the large number of firms per sector greatly simplifies the entry problem. In equilibrium, any individual entrant is negligible relative to the market, so expected profits for entrants equal the average profits of incumbent firms.

Third, the equilibrium allocation admits a closed-form solution—a distinctive feature that follows directly from the homogeneity of goods within sectors. This tractability has an important implication: computing the technology and allocative-efficiency components of TFP requires only information on the productivity distribution among active firms. The effects of other parameters—such as the elasticity of substitution—are absorbed by the endogenous productivity cutoff that determines firm activity.

This feature is crucial for decomposing TFP using standard macroeconomic data, as the empirical implementation of the model requires calibrating only the parameters governing the productivity distribution. In our empirical applications, we follow the heterogeneous-firm literature (Melitz and Redding, 2015; Edmond et al., 2015, 2023) and assume that firm productivity follows a Pareto distribution, truncated over a bounded support. Under this assumption, only three parameters must be calibrated: the lower and upper productivity bounds among active firms and the Pareto shape parameter. We discipline the shape parameter by establishing economically plausible bounds, while the productivity bounds are identified by targeting two aggregate moments: (i) aggregate TFP and (ii) the cost-weighted average markup.

To compute these target moments using real-world data, we first parameterize the production function. We assume a constant-returns-to-scale Cobb-Douglas technology with capital and labor as inputs and set the capital elasticity of output to $\alpha = 0.30$. This choice is supported by two pieces of evidence. First, this value is broadly consistent with the cost-based estimates reported by Vollrath (2024) for the United States. Second, using $\alpha = 0.30$, our cross-country estimates of the cost-weighted average markup are broadly consistent with those reported by Díez et al. (2021) for their sample of 19 countries.

Once α is determined, the two target moments can be computed directly from standard macroeconomic data. Aggregate TFP is backed out as the residual from the production function, while the cost-weighted average markup is pinned down by the ratio of the labor elasticity of output, $1 - \alpha$, to the labor income share. This markup formula is consistent with the expression derived by Hall (1988), which underlies the recent literature on firm-level markup estimation (De Loecker and Warzynski, 2012; De Loecker and Eeckhout, 2018; De Loecker et al., 2020; Traina, 2018; Calligaris et al., 2018; Autor et al., 2020; Díez et al., 2021).

Importantly, under the Pareto assumption—and, more generally, for any productivity distribution satisfying *scale invariance*—allocative efficiency can be inferred solely from the cost-weighted average markup. In particular, a higher labor share and thus a lower average markup—indicating fewer distortions—is associated with lower misallocation, with the economy converging to full efficiency as the average markup approaches its competitive, unitary level.

Once allocative efficiency is identified from the markup, the remaining component of TFP—technology—is pinned down using aggregate TFP data. In this sense, the residual from the aggregate production function corresponds not to overall TFP but to its technology component, yielding a cleaner measure of technological progress that is purged of market-power-driven misallocation effects.

Employing this quantitative strategy, we decompose aggregate TFP into technology and allocative efficiency components for up to one hundred countries over the period 1950–2019, using data from the Penn World Table (PWT) 11.0 (Feenstra et al., 2015). We use this decomposition to revisit key facts of economic growth through three empirical analyses.

First, we reexamine growth in the United States, taken as a proxy for the world income frontier. We begin by showing that our market-power-driven misallocation estimates are in line with those obtained from the oligopoly model of Edmond et al. (2023). This evidence provides an important external validation of our framework, given that their analysis relies on a more general model calibrated with firm-level data. Turning to the empirical results, we find that misallocation plays a limited role in long-run productivity growth but can be relevant at shorter horizons. In particular, allocative efficiency accounts for a substantial share of the variation in TFP growth across subperiods, whereas the growth rate of the technology component is relatively stable over time, around 1 percent per year, except for two notable episodes—1954–1973 and 2000–2007—during which technological progress accelerates to roughly 2 percent per year.

Second, we conduct a development-accounting exercise using standard success measures from the literature. We compare the conventional approach—where TFP is treated as the residual—with an alternative specification implied by our model, in which only the technology component plays that role. We find that accounting for market-power-driven misallocation increases the share of cross-country income variation explained by observables, although the majority of income differences remains unexplained.

Finally, we investigate economic convergence. The data provide little support for convergence in income or in aggregate TFP, reflecting the absence of convergence in the technology component. By contrast, we find suggestive evidence of convergence in allocative efficiency, although this pattern is not uniform across all empirical exercises. These results are consistent with the view that barriers to international technology diffusion may persist over long periods, limiting convergence in productivity and income, whereas within-country forces—such as competition or market integration—may gradually reduce market-power-driven misallocation.

Overall, these findings are robust to using alternative proxies from PWT to measure the variables of interest; to adopting different plausible calibrations for the Pareto shape parameter and the capital elasticity α; and to considering either the full sample of countries or a restricted sample excluding oil producers.

Related literature. From a methodological perspective, our work relates to a growing literature that embeds oligopolistic market structures into macroeconomic models (Bernard et al., 2003; Atkeson and Burstein, 2008; Edmond et al., 2015; Peters, 2020; De Loecker et al., 2021; Wang and Werning, 2022; Martinez and Santos, 2026; Edmond et al., 2023). Among these contributions, Edmond et al. (2023) is the closest to our paper. Their study quantifies the welfare costs of markups in the United States using an oligopoly model—calibrated with firm-level data—that incorporates channels beyond misallocation, such as inefficient entry. By contrast, our focus is on measuring exclusively the contribution of misallocation to TFP dynamics across countries and over time. On the theoretical side, our model builds on their oligopoly framework but deliberately restricts attention to a tractable special case. In particular, we consider an environment in which many firms compete within each sector while retaining strategic interaction. This structure yields several analytical simplifications. Entry becomes negligible at the sectoral level, sectors are identical in equilibrium, and the model can therefore be solved for a single representative sector, which admits a closed-form solution. Our contribution to this literature is to develop a framework that can be calibrated using standard macroeconomic aggregates rather than detailed firm-level data. This feature allows us to quantify market-power-driven misallocation across a broader set of countries and years.

Given our goal of revisiting key facts of economic growth, this study is closely related to the development economics literature, particularly the strand focused on macroeconomic analyses and development accounting exercises (Klenow and Rodríguez-Clare, 1997; Prescott, 1998; Caselli, 2005, 2016; Jones, 2016). Jones (2016) serves as our primary reference, offering a comprehensive overview of economic growth facts, some of which we reexamine here. Our contribution to this literature lies in providing a systematic assessment of misallocation across both time and space. In particular, we incorporate allocative efficiency into development accounting exercises over multiple years. Furthermore, we provide evidence on the effects 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). Our approach is most closely related to Baqaee and Farhi (2020), who also measure allocative efficiency as the distance from the optimal allocation.[1] Their framework is considerably more general, allowing for arbitrary elasticities of substitution, returns to scale, factor mobility, and input-output network linkages, without imposing a specific market structure. As a consequence, its empirical implementation is highly data-intensive, requiring detailed firm-level information, whereas our framework can

[1] Measuring allocative efficiency as the distance from the 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), "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."

be implemented using standard macroeconomic data.

Finally, our paper is also related to the literature on misallocation (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009; Restuccia and Rogerson, 2013). In this literature, misallocation typically arises from exogenous wedges, whose quantification requires extensive firm-level data, much like in growth-accounting approaches—particularly Baqaee and Farhi (2020). By contrast, in our framework misallocation emerges endogenously and can be quantifiable using macroeconomic aggregates. Moreover, whereas the existing literature typically focus on explaining cross-country TFP differences—often within the manufacturing sector due to data constraints—we analyze economy-wide outcomes across both across countries and time.

The remainder of the paper proceeds as follows. Section 2 presents the model. Section 3 outlines the quantitative strategy, and Section 4 describes the data used in the analysis. Section 5 examines productivity dynamics at the global income frontier, proxied by the United States. Section 6 examines cross-country income differences through a development-accounting perspective, while Section 7 investigates economic convergence. Section 8 presents a set of robustness checks, and Section 9 briefly discusses a model extension. Finally, Section 10 concludes.

2 Model

The economy is closed and features a representative consumer with preferences over final consumption, who supplies labor inelastically and owns all firms. The final good is produced by perfectly competitive firms using highly differentiated inputs from a continuum of sectors. Within each sector, a large number of heterogeneous-productivity firms produce a homogeneous good using capital and labor and engage in Cournot competition. There are no fixed operating costs. Firms enter by paying a sunk cost denominated in monetary terms—assumed to be low—after which they receive a productivity draw and are randomly allocated to a sector. All firms exit at the end of the period, rendering firm decisions static.

The remainder of this section lays out the model in detail, focusing on its economic structure. Formal proofs and derivations are collected in Appendix A.

Households. The model features a representative consumer with preferences over final consumption, who owns all firms and supplies labor inelastically. Since capital takes time to build, both capital and labor are predetermined at the beginning of each period. This assumption simplifies the entry problem and allows us to focus on the firm side, where misallocation originates.[2] Moreover, it aligns with the concept of misallocation, which concerns the maximum output attainable *given* the total amount of inputs available.

[2]More elaborate household specifications would merely affect the entry decision, altering the precise interpretation of our low-barriers assumption. The main results—such as the expression for misallocation—would remain unaffected.

Final good producers. Perfectly competitive firms produce a homogeneous final consumption good Y using inputs Y_s from a continuum of sectors through the CES production function

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

where $\eta > 0$ denotes the elasticity of substitution across sectors $s \in [0, 1]$.[3] We assume that η is low—that is, products are highly differentiated across sectors. With the final good taken as the numeraire, the first-order conditions (FOCs) of the representative final-good producer's profit-maximization problem are

$$p_s = (Y/Y_s)^{1/\eta}, \quad \forall s \in [0, 1]. \tag{2}$$

where p_s represents the price of goods within sector s, which are homogeneous across all firms in that sector. Moreover, the zero-profit condition implies

$$1 = \left(\int_0^1 p_s^{1-\eta} ds \right)^{\frac{1}{1-\eta}}. \tag{3}$$

Intermediate goods producers. Within each sector, a continuum of firms produces a homogeneous good:

$$Y_s = \int_0^{N_s} Y_{si} di. \tag{4}$$

Firm i in sector s has productivity $A_{si} > 0$ and produces according to

$$Y_{si} = A_{si} K_{si}^{\alpha} L_{si}^{1-\alpha}, \tag{5}$$

where $K_{si} \geq 0$ is the stock of capital, $L_{si} \geq 0$ is the labor employed, and $\alpha \in (0, 1)$. Productivity is drawn from a distribution with density $g(A)$ and cumulative probability $G(A)$, defined over the bounded support $[\underline{\underline{A}}, \overline{A}] \subset (0, \infty)$, where $\overline{A}$ represents the technology frontier. There are no operating fixed costs.

Intermediate firms engage in Cournot competition, choosing their output taking as given the output chosen by the other firms in the sector, as well as the wage $w > 0$ and the rental cost of capital $r > 0$. They face the inverse demand function (2). Since each firm is negligible in the aggregate economy, it takes final output and the final-good price as given. The FOC for an intermediate firm is

$$p_s = MC_{si} \frac{\eta_{si}}{\eta_{si} - 1}, \tag{6}$$

[3] Given that firms' decisions are static, we omit time subscripts for notational simplicity.

where $MC_{si} = \left(\frac{r}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\frac{1}{A_{si}}$ is the Cobb-Douglas marginal cost of firm i from sector s, while $\eta_{si} = \frac{\eta}{s_{si}}$ is the price elasticity of demand faced by a firm with sectoral market share $s_{si} \equiv Y_{si}/Y_s$. Thus, the firm's markup is $\mu_{si} \equiv \frac{\eta_{si}}{\eta_{si}-1} = \left(1 - \frac{s_{si}}{\eta}\right)^{-1}$.[4]

Assessing the second-order condition (SOC) requires knowledge of the firm's market share, which is determined only after solving the equilibrium allocation. We initially assume that the SOC holds and subsequently verify it using the equilibrium expression for s_{si}. The condition is satisfied for all $\eta \in (0, 1]$, and also for $\eta > 1$, provided it is not too large.[5] Hence, under the low-elasticity assumption maintained throughout, Equation (6) represents firm i's optimal decision for all active firms.

Equation (6) further implies that more productive active firms must have larger market shares. Since a more productive firm has lower marginal cost, it charges a higher markup—which can only occur if its market share is larger.

Equilibrium allocation. Since sectors are ex ante equivalent and each contains a large number of firms—both total and active—they are also ex post equivalent. As a result, in equilibrium, the productivity distribution is identical across sectors, each represented by the same density function g introduced earlier. Moreover, from Equations (1) and (3), output and prices satisfy $Y = Y_s$ and $p_s = 1$ for any sector s. The same logic applies to all other equilibrium variables. Accordingly, we drop the sectoral subscript for the remainder of this section.

In contrast to models with heterogeneous goods, when goods are homogeneous within sectors, some firms may remain inactive even in the absence of fixed costs. Because prices are equalized within sectors, firms with high marginal costs may not operate profitably. This selection becomes increasingly stringent as the number of firms rises: competition implies that only sufficiently productive firms can remain active. In particular, when total entry exceeds a threshold $\check{N}$, not all firms can be simultaneously active.[6] We focus on the case of low entry barriers, so that in equilibrium $N > \check{N}$ and some low-productivity firms remain inactive. In that case, let $\underline{A} > \underline{\underline{A}}$ denote the endogenous productivity cutoff for active firms such that a firm is active if and only if $A_i \geq \underline{A}$.[7] At the cutoff, the marginal active firm produces zero output, so $s(\underline{A}) = 0$ and profits are zero.

We can now solve for the equilibrium allocation. Combining Equation (6), price equaliza-

[4]This markup expression is standard and remains valid with either a discrete or continuous set of firms because the only requirement from the sectoral aggregator—Equation (4) in our model—is the marginal contribution $\partial Y_s/\partial Y_{si}$, which equals one in both cases.

[5]Formally, by Proposition A.4, the SOC holds if $\eta \in \left(0, \frac{1+\underline{\underline{A}}/\overline{A}}{1-\underline{\underline{A}}/\overline{A}}\right]$.

[6]See Equation (A.3) in Appendix A.1 for a formal definition of $\check{N}$.

[7]This result is consistent with the oligopoly models of Atkeson and Burstein (2008), Edmond et al. (2015), and De Loecker et al. (2021), where only a few firms compete in each sector. In these models, the equilibrium set of active firms is generally not unique—a multiplicity eliminated by ruling out equilibria in which an inactive firm has a lower marginal cost than an active one. This implies that a firm i is active if $A_i > \underline{A}$ and only if $A_i \geq \underline{A}$, since not all firms with productivity exactly $\underline{A}$ need be active. However, such a situation does not arise with a continuum of firms, yielding the stated condition.

tion within sectors, and the condition $s(\underline{A}) = 0$, we obtain

$$s(A_i) \equiv s_i = \eta \left(1 - \underline{A}/A_i\right) \tag{7}$$

$$\eta = \frac{1}{N_a \left[1 - \mathrm{E}_a \left(\underline{A}/A\right)\right]}, \tag{8}$$

where $\mathrm{E}_a \left(h(A)\right) \equiv \mathrm{E} \left(h(A) \mid A \geq \underline{A}\right) = \int_{\underline{A}}^{\overline{A}} h(A) \frac{g(A)}{1-G(\underline{A})} dA$ denotes expectations over active firms, and $N_a \equiv N \left(1 - G(\underline{A})\right)$ is the mass of active firms.

The equilibrium allocation therefore admits a closed-form solution. This distinctive feature of our model follows directly from the homogeneity of goods within sectors. With heterogeneous goods, price equalization no longer holds, and relative prices would be a nonlinear function of relative market shares, precluding an analytical solution.

Equation (7) implies that more productive firms have larger market shares and earn higher profits, while the marginal firm with productivity $\underline{A}$ produces no output and earns zero profit. Equation (8) implicitly determines the cutoff $\underline{A}$ as a function of total entry N. In particular, $\underline{A}$ is increasing in N for $N \geq \check{N}$: greater entry tightens selection and raises the productivity threshold. When $N\eta \to \infty$, productivity dispersion among active firms vanishes as $\underline{A} \to \overline{A}$, and $N_a \to \infty$. Moreover, when η is low, the mass of active firms N_a must be large, since $\mathrm{E}_a \left(\underline{A}/A\right)$ lies in $[0, 1]$. This provides an economic rationale for modeling a continuum of firms within each sector: weak competitive pressure across sectors (low η) sustains a large number of heterogeneous active firms.

Aggregate productivity and misallocation. Since intermediate firms face the same factor prices, optimal input choice implies that they all select the same capital–labor ratio. Let aggregate capital and labor be defined as $K \equiv \int_0^N K_i di$ and $L \equiv \int_0^N L_i di$. It follows that $K_i/K = L_i/L \equiv \theta_i$ for every active firm i. Combining this result with Equations (4), (5), (7), and (8), aggregate output can be written as

$$Y = \overline{A} \Omega K^{\alpha} L^{1-\alpha}, \tag{9}$$

where

$$\Omega = \frac{\mathrm{E}_a \left[(\underline{A}/\overline{A})(1 - \underline{A}/A)\right]}{\mathrm{E}_a \left[(\underline{A}/A)(1 - \underline{A}/A)\right]}. \tag{10}$$

Since goods are homogeneous within sectors and firms have constant returns to scale, the efficient allocation assigns all inputs to the most productive firm. In this case, $\Omega = 1$, and total output is $Y = \overline{A} K^{\alpha} L^{1-\alpha}$. Consequently, $\Omega \in (0, 1]$ represents the wedge between actual TFP, $\overline{A}\Omega$, and its efficient level, $\overline{A}$, and thus serves as our measure of allocative efficiency. Notably, Ω increases in the cutoff $\underline{A}$: as the productivity threshold rises and dispersion among active firms shrinks, allocative efficiency improves. In the limit as $\underline{A} \to \overline{A}$, productivity dispersion vanishes

and the allocation becomes fully efficient ($\Omega \to 1$).

Markups. Substituting Equation (7) into $\mu_i = \left(1 - \frac{s_i}{\eta}\right)^{-1}$ yields

$$\mu(A_i) \equiv \mu_i = A_i/\underline{A}, \tag{11}$$

so a firm's markup depends solely by its productivity relative to the marginal firm. The marginal firm hence charges a unit markup, whereas more productive firms optimally set higher markups.

Equation (11) also implies that the markup distribution is fully determined by the productivity distribution among active firms. Consequently, knowledge of this distribution is sufficient to compute aggregate markup measures. In particular, consider the cost-weighted average markups, defined as $\mu \equiv \int_0^N \left(\frac{L_i w + K_i r}{Lw + Kr}\right) \mu_i di = \int_0^N \theta_i \mu_i di$, where the second equality holds because every active firm i chooses the same capital–labor ratio, implying $K_i/K = L_i/L \equiv \theta_i$. As a result,

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

Edmond et al. (2023) argue that the cost-weighted average markup is the most appropriate measure of the overall level of markups, as it captures the relevant wedge in aggregate employment and investment decisions.[8] As we discuss in Section 3, this expression plays a central role in our quantitative strategy.

Entry decision and free-entry equilibrium. Firms enter by paying a sunk cost $f_e > 0$ denominated in monetary terms. Upon entry, a firm draws productivity $A_i \sim G(A)$ and is randomly assigned to a sector $s \in [0, 1]$. This cost can be interpreted as a licensing fee, potentially rebated to households, though it does not involve the use of real resources (e.g., labor or capital) and therefore plays no further role in the model. If instead the entry cost were defined in units of labor, an increase in entry would reduce the labor available for production. Since only labor actually used in production should enter the aggregate production function, such a formulation is at odds with standard measures of TFP used in growth and development accounting. By adopting a monetary entry cost instead, we preserve standard TFP measurement, allowing us to focus on the role of misallocation in shaping productivity.

The entry decision depends on the expected profit of an entrant. Since all firms exit at the end of each period, it suffices to compute expected profit over a single period. With a large number of firms the expected profit of an entrant is simply equal to the average profit of incumbent firms, $\bar{\pi}$, as the entry of any single firm does not affect aggregate outcomes. Recalling that

[8] Edmond et al. (2015) and Baqaee and Farhi (2020) instead rely on the sales-weighted *harmonic* average markup. However, as Edmond et al. (2023) point out, when firms share the same returns to scale, these two measures of aggregate markup coincide. Appendix A.6 provides a formal proof of this equivalence in our model.

$p_s = 1$ and $Y_s = Y$ for all sectors s, we obtain:

$$\bar{\pi} = K^{\alpha} L^{1-\alpha} \left(\frac{\overline{A}\Omega - \underline{A}}{N} \right), \tag{13}$$

which follows from Equations (7), (8), (9), (10), and (11).

In the free-entry equilibrium, the condition $f_e = \bar{\pi}$ must hold, implicitly defining the total number of firms as a decreasing function of the entry cost, $N(f_e)$. This result follows directly when $\bar{\pi}$ decreases with N, but it continues to hold even if $\bar{\pi}$ is non-monotonic—a theoretical possibility in our model arising under extreme scenarios. For simplicity, we focus on the monotonic case in the main text and relegate the general case to Appendix A.8.

Therefore, a large number of entrants—as implied by our assumption of low barriers to entry—requires a small f_e. More precisely, $N > \check{N}$ if $f_e < \check{f}_e$, where $\check{f}_e$ denotes the average profit of incumbent firms when $N = \check{N}$.[9]

Naturally, the specific value of f_e within its domain $(0, \check{f}_e)$ plays a critical role: a lower sunk cost induces greater entry, raising the cutoff $\underline{A}$ and thereby reducing misallocation, as markups move closer to one (Equation (11)). In the limit, as $f_e \to 0$, we have $\underline{A} \to \overline{A}$ and $\Omega \to 1$, as entry becomes arbitrarily large.

From oligopoly to perfect competition. From the preceding analysis, when barriers to entry are low—i.e., $f_e \in (0, \check{f}_e)$—the number of firms entering the market is large ($N > \check{N}$), so some low-productivity firms remain inactive ($\underline{A} > \underline{\underline{A}}$). The number of firms is large not only in the sense that not all firms can be simultaneously active; it is also large in absolute terms, allowing us to represent the distribution of firms using continuous functions. This latter feature is ensured by the low elasticity of substitution η, as Equation (8) makes clear.[10] Taken together, these results imply that a large number of firms per sector arises from two ingredients: a low entry cost f_e and a low elasticity of substitution η.

A large number of firms, however, does not by itself eliminate strategic interaction. Market power vanishes only when N becomes large *relative* to $1/\eta$—that is, as $\eta N \to +\infty$.[11] This situation arises when $f_e \to 0$, in which case $\underline{A} \to \overline{A}$, $N_a \to +\infty$, $\mu_i \to 1$ for all active firms i, and $\Omega \to 1$. In words, this limiting case corresponds to perfect competition, with infinitely many active firms all having the same productivity, unit markups, and zero profits—and no misallocation.

[9] See Equation (A.12) in Appendix A.8 for the formal definition of $\check{f}_e$, derived from Equation (13).

[10] These two notions of "large" become compatible because $\check{N}$ itself is large when η is low. See Equation (A.3) in Appendix A.1.

[11] This observation aligns with Equation (6), which shows that a firm with a markup close to one must have a market share that is small relative to the elasticity of substitution η. With low elasticity, even a market share close to zero can generate a high markup.

3 Quantitative strategy for TFP decomposition

This section outlines the quantitative strategy used to implement the model for decomposing TFP, for a given country and period, into its two components: technology, $\overline{A}$, and allocative efficiency, Ω. From Equation (10), allocative efficiency depends solely on the productivity distribution among active firms. In oligopoly settings, the aggregate misallocation wedge is governed by the distribution of markups, which in our model can be inferred directly from productivity through Equation (11). Crucially, the effects of other parameters—such as the elasticity of substitution η or the entry cost f_e—are absorbed by the endogenous productivity cutoff $\underline{A}$ (see Equation (8)). This distinctive property follows from the model's assumptions—most notably, low barriers to entry, homogeneous goods, and the absence of fixed costs—which jointly imply that one market share is known independently of all parameter values: the marginal active firm always produces zero output.

As a result, computing misallocation does not require calibrating the full set of structural parameters. Instead, it suffices to characterize the productivity distribution, which we assume to be Pareto in the empirical applications. We propose a calibration procedure that identifies the productivity bounds among active firms, $\underline{A}$ and $\overline{A}$, by targeting two moments: (i) aggregate TFP, $\overline{A}\Omega$, and (ii) the cost-weighted average markup, μ.

After presenting the calibration strategy, we describe how the two target moments are computed using standard macroeconomic data together with the capital elasticity α. We then detail our choices for α and for the remaining parameter governing the productivity distribution, which determines the shape of the Pareto density.

3.1 Calibrating the productivity bounds among active firms

We calibrate the productivity bounds $\underline{A}$ and $\overline{A}$, for each period and country, by targeting two moments: (i) aggregate TFP, $\overline{A}\Omega$, and (ii) the cost-weighted average markup, μ. Formally, the calibration proceeds as follows:

1. From Equation (12), compute the lower bound as $\underline{A} = \overline{A}\Omega/\mu$.

2. Given $\underline{A}$ from the previous step, recover the upper bound $\overline{A}$ by 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 follows directly from Equations (10) and (12).

In the theoretical analysis of Section 2, the upper bound $\overline{A}$ is treated as exogenous, and the focus lies on solving for the endogenous cutoff $\underline{A}$. Empirically, however, the situation is reversed. The two target moments pin down the lower bound directly via Equation (12) (step 1), while the main challenge is to recover $\overline{A}$ by solving Equation (14) (step 2).

Evaluating Equation (14) requires only moments of the productivity distribution among active firms, which depend on the assumed functional form. Following much of the heterogeneous-firm literature (Melitz and Redding, 2015; Edmond et al., 2015, 2023), we assume that productivity follows a Pareto distribution with shape parameter $k \neq 0$, truncated over the interval $A \in \left[\underline{A}, \overline{A}\right]$. The associated density is increasing in A for $k < -1$, constant for $k = -1$ (i.e., equivalent to the uniform distribution), and decreasing for $k > -1$, $k \neq 0$. For now, we treat k as given and return to its calibration in Section 3.3.

Appendix B characterizes the conditions under which Equation (14) admits a solution. In particular, Proposition B.8 shows that, under a Pareto distribution, a unique solution exists for all μ satisfying (i) $\mu > 1$ and (ii) $k < 2/(\mu - 1)$. If $\mu > 1$ but $k \geq 2/(\mu - 1)$—a case that can arise only when $k > 0$—no solution exists. Accordingly, in the empirical applications, country-period observations for which these conditions are violated are excluded, as the model does not admit a well-defined solution in those cases. Similar existence and uniqueness results can be derived for other productivity distributions.[12]

A distinctive feature of the truncated Pareto distribution is that it satisfies *scale invariance*: multiplying all productivity levels by a positive scalar leaves the distribution unchanged. Although this property is not technically required for the calibration algorithm to be well defined, it is highly desirable in practice, since the level of TFP—and thus the absolute productivity level of all firms—is essentially arbitrary.

Scale invariance has an important implication. The productivity moments appearing in Equation (14) depend not on the absolute values of the productivity bounds, but only on their ratio, $\tilde{A} \equiv \overline{A}/\underline{A} > 1$, together with the shape parameter k.[13] This allows step 2 of the calibration algorithm to be implemented first by solving for $\tilde{A}$, which in turn delivers allocative efficiency Ω via Equation (10). Hence, estimating misallocation requires only data on the average markup μ. TFP data—whose level is arbitrary—does not affect Ω, but determines absolute productivity levels by enabling the separate identification of $\underline{A}$ and $\overline{A}$. In particular, once Ω is known, the upper bound $\overline{A}$ follows directly from observed TFP, $\overline{A}\Omega$. In this sense, in our model, the residual from the production function is not TFP itself, but only its technology component, $\overline{A}$—a cleaner residual, as it is purged of misallocation effects.

The preceding reasoning implies that allocative efficiency Ω can be expressed as a function of the average markup μ. Figure 1 illustrates this relationship for truncated Pareto distributions with $k = 3, 5, 9$, as well as for the uniform distribution ($k = -1$). Two main features stand out. First, Ω is decreasing in μ, with $\Omega \to 1^-$ as $\mu \to 1^+$. Essentially, this function maps one measure of distance from the efficient equilibrium into another. While μ captures the deviation of the average markup from its competitive benchmark, $\mu = 1$, allocative efficiency Ω reflects

[12] Proposition B.4 establishes that, for *any* continuous truncated distribution defined for all $\overline{A} \in (\underline{A}, +\infty)$, a unique solution exists for all $\mu \in \left(1, \lim_{\overline{A}\to+\infty} \mu\right)$, and no solution exists for higher average markups. In other words, $\overline{A}$ can be recovered provided that μ lies above one and below an upper bound, which may itself be infinite and depends on the assumed productivity distribution.

[13] Closed-form expressions for these moments under a Pareto distribution are derived in Proposition B.5.

the deviation of aggregate TFP, $\overline{A}\Omega$, from its efficient level, $\overline{A}$. Economies with markups closer to unity therefore exhibit higher allocative efficiency.

Second, the mapping between Ω and μ depends on the Pareto shape parameter k. Lower values of k (i) map a given markup μ into a higher level of allocative efficiency Ω and (ii) make Ω less sensitive to variations in μ.[14]

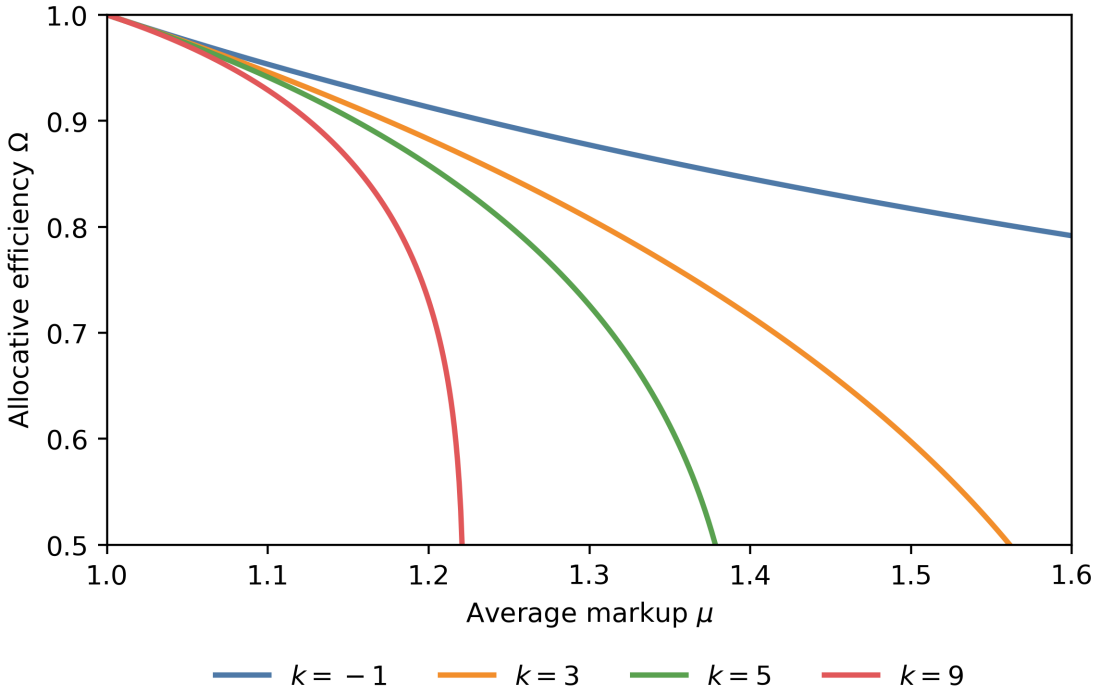


Figure 1: Allocative efficiency as a function of the cost-weighted average markup.

3.2 Computing the target moments using real-world data

Both target moments can be computed using standard macroeconomic data together with the capital elasticity α. Aggregate TFP is backed out as a residual from the aggregate production function (9): $\overline{A}\Omega = \frac{Y}{K^{\alpha}H^{1-\alpha}}$.

The cost-weighted average markup μ is pinned down by the labor income share, defined as $LS \equiv \frac{Lw}{Yp}$. Formally, since $\mu \equiv \int_0^N \left(\frac{L_i w + K_i r}{Lw + Kr}\right) \mu_i di = \int_0^N \theta_i \mu_i di$ and $\theta_i = L_i/L$, we obtain

$$\mu = \int_0^N \left(\frac{L_i}{L}\right)\left(\frac{p}{w/MPL_i}\right) di = \frac{1-\alpha}{LS}, \tag{15}$$

where MPL_i denotes the marginal product of labor. At the firm level, this implies $\mu_i = \frac{p}{w/MPL_i} = \frac{1-\alpha}{LS_i}$, with $LS_i \equiv \frac{L_i w}{Y_i p}$ representing the firm-level labor share, which corresponds to the Hall (1988) formula that underlies much of the recent literature on markup estimation

[14] In the limit, as $k \to -\infty$, allocative efficiency converges to its efficient benchmark, $\Omega \to 1$, independently of the average markup μ, as shown in Proposition B.9.

(De Loecker and Warzynski, 2012; De Loecker and Eeckhout, 2018; De Loecker et al., 2020; Traina, 2018; Calligaris et al., 2018; Autor et al., 2020; Díez et al., 2021).

Unlike this literature, which relies on firm-level estimates to construct aggregate markup measures, Equation (15) allows the cost-weighted average markup to be computed directly from macroeconomic data. This is possible because, when markups are weighted by the cost of the variable input used to compute firm-level markups, the resulting average markup depends only on aggregates (De Loecker et al., 2020). This simplification does not extend to other aggregate measures, such as the sales-weighted average markup. Importantly, the markup estimates implied by Equation (15) are consistent with aggregate movements in the labor share by construction, addressing concerns that micro-based markup measures may be difficult to reconcile with macroeconomic evidence (Basu, 2019).

3.3 Parameter choices

The final step of our quantification strategy is to discuss the parameter choices for (i) the capital elasticity α, which is required to compute the target moments, and (ii) the Pareto shape parameter k, which enters the productivity moments appearing in Equation (14).

Capital elasticity. We set the capital elasticity to $\alpha = 0.30$ in the baseline calibration. This choice is motivated by two pieces of evidence.

First, in our model, cost minimization implies that α equals capital's share in total factor costs, that is, $\alpha = \frac{Kr}{Kr+Lw}$. Implementing this identity requires estimating capital expenses, Kr. A widely used approach, including in recent work such as Barkai (2020), relies on the user-cost formula of Hall and Jorgenson (1967) to compute the required rate of return r. Using this methodology, Vollrath (2024) reports estimates for the US over the period 1948–2018.[15] On average, his baseline user-cost-based estimates imply values of α around 0.28–0.29 (see Table 2, Panel A, in Vollrath, 2024). Given the uncertainty surrounding capital expense estimates, Vollrath (2024) also proposes plausible bounds for α in the US, based on extreme assumptions about capital costs. Using data from PWT 11.0, we construct analogous bounds for the global economy based on aggregate cost shares (Appendix D). The results indicate that our baseline choice $\alpha = 0.30$ is plausible, as it lies comfortably within the global bounds (Figure D.1).

Second, building on De Loecker and Warzynski (2012) and De Loecker et al. (2020), Díez et al. (2021) estimate firm-level markups between 2000 and 2015 for 19 advanced and emerging economies, accounting for roughly 70 percent of world GDP.[16] Their analysis relies on the Orbis dataset, which covers both privately held and publicly listed firms. Applying Equation (15) at

[15] In models with input-output linkages, aggregate production elasticities generally depend on the production network and cannot be recovered from aggregate factor cost shares alone (Baqaee and Farhi, 2020). Accordingly, the baseline estimates in Vollrath (2024) are obtained by computing capital expenses at the industry level and then using input-output information to recover aggregate elasticities.

[16] The 19 countries are Belgium, Bulgaria, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Great Britain, Greece, Italy, Japan, Korea, Latvia, Portugal, Romania, Russia, Spain, and the United States.

the group level using PWT 11.0 labor share data, we find that $\alpha \approx 0.31$ matches our markup series for these 19 countries with the cost-weighted average markup series reported in Figure 1 of Díez et al. (2021).[17]

Figure 2 illustrates the implications of our baseline choice $\alpha = 0.30$ for cross-country cost-weighted average markups. Overall, our estimates are broadly consistent with the series reported by Díez et al. (2021) for their 19-country sample. While the two series do not coincide year by year, they deliver the same qualitative conclusion of a mild upward trend over 2000–2015. Moreover, our estimates for the same 19 countries are only slightly higher on average, consistent with our baseline choice $\alpha = 0.30$, which is marginally below the value that best matches their series ($\alpha \approx 0.31$). Turning to the broader non-oil sample, the overall trend remains roughly similar, but the level is about 5 percent higher, suggesting that global average markups may exceed those reported by Díez et al. (2021) for their 19-country sample.

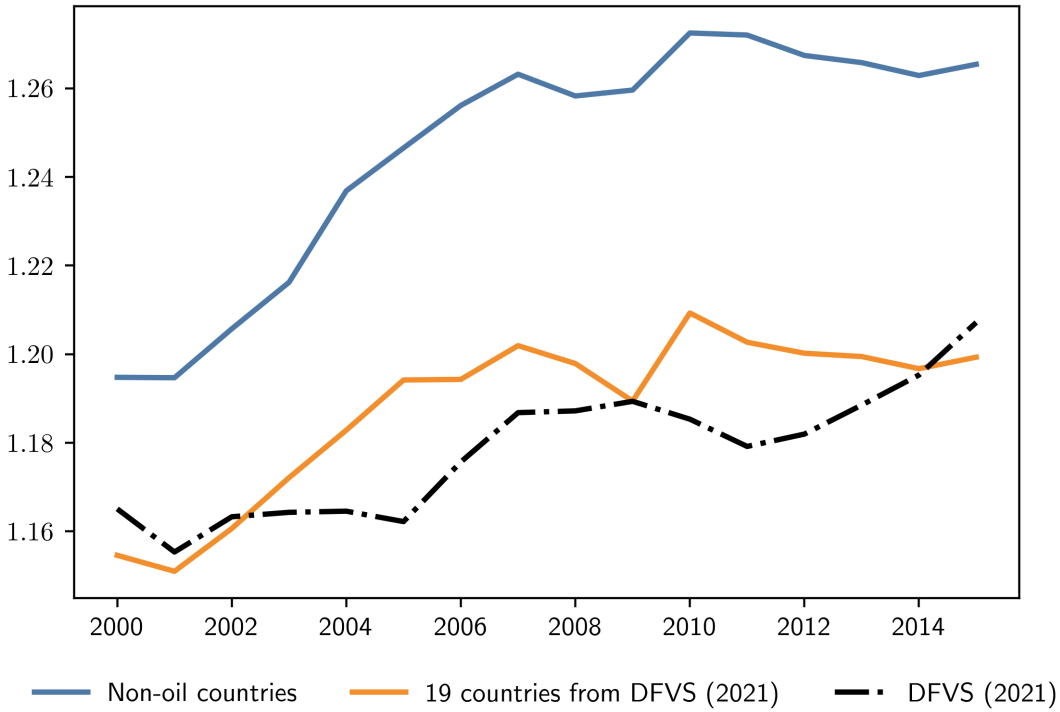


Figure 2: Cross-country cost-weighted average markups, 2000–2015. *Notes*: the blue and red lines report our estimates of the cost-weighted average markup, computed by applying Equation (15) at the group level using (i) the baseline $\alpha = 0.30$ and (ii) GDP-weighted average labor shares across countries, based on PWT 11.0 data (variables *labsh* and *cgdpo*). The two series differ only in sample coverage: the blue line includes all available non-oil countries, while the red line restricts the sample to the 19 economies in the baseline sample of Díez et al. (2021). The black dashed line reproduces the cost-weighted average markup for this 19-country sample reported in Figure 1 of Díez et al. (2021).

[17] Group-level labor shares are computed as GDP-weighted averages across countries. Country-level labor shares are measured using the PWT *labsh* variable, while GDP (in current PPPs) is measured using *cgdpo*.

Pareto shape parameter. In the absence of a natural point estimate for the Pareto shape parameter k, we rely on economically plausible bounds. As shown in Figure 1, lower values of k make allocative efficiency Ω less sensitive to variations in the average markup μ. As a result, when k is low, misallocation varies little across countries or over time, regardless of the observed heterogeneity in markups. Conversely, higher values of k allow for greater variation in allocative efficiency in response to markup differences. In this sense, the resulting bounds on k characterize a plausible range for the potential role of misallocation in explaining TFP differences.

As shown in Section 3.1, under a Pareto productivity distribution a necessary condition for the calibration algorithm to exactly match the two target moments is $k < 2/(\mu - 1)$. This restriction provides a natural upper bound for k, which depends on the magnitude of the cost-weighted average markup μ. Figure 3 presents box plots of the empirical markup distribution across non-oil countries. These country-level average markups are computed by applying Equation (15), using PWT 11.0 labor share data and the baseline $\alpha = 0.30$. The figure indicates that a markup of $\mu \approx 1.8$ lies near the upper tail of the distribution—above the 90th percentile for most of the sample and close to it in recent years. This value implies $k < 2/(1.8 - 1) \approx 2.5$. To remain conservative, we set $k = 3$ as the upper-bound value in our calibration.

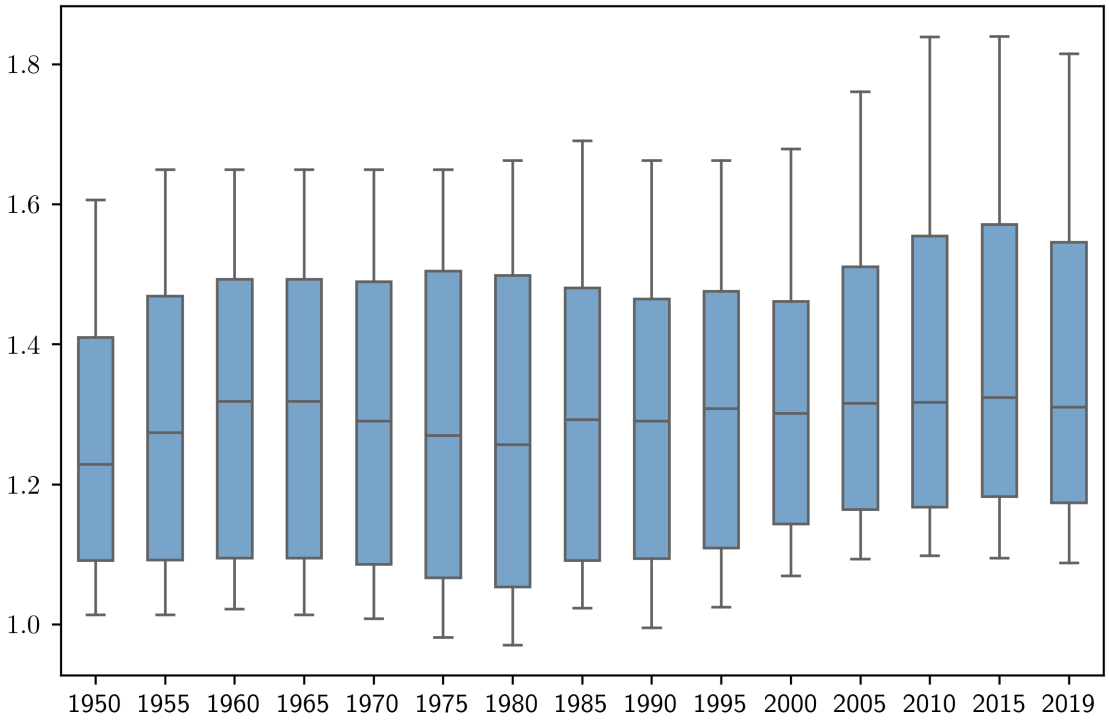


Figure 3: Distribution of cost-weighted average markups μ across non-oil countries. *Notes*: for each country and year, the average markup μ is computed from Equation (15), using the baseline $\alpha = 0.30$ and PWT 11.0 labor share data (*labsh* variable). In each year, boxes span the interquartile range (25th–75th percentiles), the horizontal line denotes the median, and whiskers extend to the 10th and 90th percentiles. Outliers are not shown.

In contrast, there is no strict mathematical lower bound on k: for any $k < 0$, the cali-

bration algorithm admits a unique solution as long as $\mu > 1$. The relevant restriction in this case is economic rather than technical. Values of $k < -1$ imply a productivity density that is upward-sloping, meaning that highly productive firms are relatively abundant. Such a distribution appears implausible in most empirical settings and contrasts with the common assumption that highly productive firms are scarce, which is often imposed by assuming that productivity is drawn from a non-truncated Pareto distribution. We therefore set the lower bound at $k = -1$, which corresponds to a uniform productivity distribution.

Together, these arguments imply that the empirically relevant range for the Pareto shape parameter is $k \in [-1, 3]$. In the quantitative analysis, we therefore focus on the lower and upper bounds of this interval, which provide a plausible range for the potential contribution of misallocation.

4 Data

Our analyses focus on non-oil countries to avoid contamination of our measures of average markups and allocative efficiency by natural resource rents.[18] A country is classified as an oil producer in a given year if its oil rents account for more than 10 percent of GDP. Data on oil rents as a share of GDP are obtained from the World Development Indicators (WDI), provided by the World Bank.

For all other variables, we use data from the Penn World Table (PWT) 11.0, which contains annual information for 185 countries from 1950 to 2023. To avoid the pandemic period, we restrict the analysis to 1950–2019. Our empirical exercises include both time-series (growth) analyses and cross-country (level) comparisons. As emphasized by Feenstra et al. (2015), national-accounts-based variables are best suited for within-country growth analyses, whereas variables expressed in current PPPs are preferred for cross-country level comparisons. Accordingly, we construct separate datasets for growth and level analyses. To balance the trade-off between using the most accurate variable measures and maximizing sample coverage, we construct two datasets for each type of analysis: one using the most accurate available proxies but covering fewer country-year observations, and another with broader coverage but based on less precise measures.

Altogether, this strategy yields four distinct datasets, summarized in Table 1. Using PWT data, we measure output (GDP) Y, GDP per capita, capital K, labor L, and the labor income share LS. We also consider raw employment, denoted by E, measured either by the number of persons engaged (to maximize coverage) or by total hours worked (when using the best proxies). Labor L—the input that enters the production function—is defined as raw employment adjusted for human capital, based on years of schooling and returns to education.

[18] In oil-producing economies, a substantial share of value added reflects rents from natural resources rather than returns to reproducible capital or labor. As a result, labor income shares tend to be mechanically low (Feenstra et al., 2015), which would translate into inflated average markup estimates via $\mu = (1 - \alpha)/LS$ (Equation (15)).

Table 1: Variable definitions across datasets

	Datasets for level analyses		Datasets for growth analyses	
	Broad coverage	Best proxies	Broad coverage	Best proxies
Output, Y	Output-side GDP, PPP (PWT, *cgdpo*)	Output-side GDP, PPP (PWT, *cgdpo*)	GDP, NA (PWT, *rgdpna*)	GDP, NA (PWT, *rgdpna*)
GDP per capita	GDP per capita, PPP (PWT, *cgdpo*/*pop*)	GDP per capita, PPP (PWT, *cgdpo*/*pop*)	GDP per capita, NA (PWT, *rgdpna*/*pop*)	GDP per capita, NA (PWT, *rgdpna*/*pop*)
Capital, K	Capital stock, PPP (PWT, *cn*)	Capital services, PPP (PWT, *ck*)	Capital stock, NA (PWT, *rnna*)	Capital services, NA (PWT, *rkna*)
Raw employment, E	Persons engaged (PWT, *emp*)	Persons engaged × hours (PWT, *emp*×*avh*)	Persons engaged (PWT, *emp*)	Persons engaged × hours (PWT, *emp*×*avh*)
Labor, L	E × human capital (PWT, *emp*×*hc*)	E × human capital (PWT, *emp*×*avh*×*hc*)	E × human capital (PWT, *emp*×*hc*)	E × human capital (PWT, *emp*×*avh*×*hc*)
Labor share, LS	Baseline labor share (PWT, *labsh*)	Preferred labor share (PWT, *lab_sh2*)	Constructed series (see Appendix E)	Preferred labor share (PWT, *lab_sh2*, no imput.)
Oil rents / GDP	Oil rents / GDP (WDI)	Oil rents / GDP (WDI)	Oil rents / GDP (WDI)	Oil rents / GDP (WDI)

Notes: PPP denotes current purchasing power parities. NA denotes constant national prices. "No imput." indicates that interpolated and extrapolated observations are excluded.

As shown in Table 1, the datasets based on the best proxies differ from their broad-coverage counterparts along three dimensions. First, capital K is measured using capital services rather than capital stocks. Second, raw employment E and labor L are adjusted for differences in hours worked. Third, the labor income share LS is measured using the preferred PWT series (*lab_sh2*), which assumes that self-employed workers use labor and capital in the same proportions as the rest of the economy (Feenstra et al., 2015).

In the broad-coverage datasets, we instead rely on the baseline PWT labor share series (*labsh*) for level analyses and construct an alternative series for growth analyses. This new series is better suited for measuring labor share growth, in part because it excludes interpolated and extrapolated observations in PWT data, which can mechanically distort measured growth rates (see Appendix E for full details). Accordingly, we also exclude such imputed observations from the preferred PWT labor share series used in the best-proxy growth dataset.

Figure 4 reports the number of non-oil countries for which TFP can be decomposed, by dataset, value of k, and year, over the 1950–2019 period. As discussed in Section 3.1, we retain only country-year observations for which the cost-weighted average markup $\mu = (1-\alpha)/LS$ satisfies $\mu > 1$ and $k < 2/(\mu - 1)$, as the model does not admit a well-defined solution otherwise. Several patterns emerge. First, the use of interpolated and extrapolated labor share observations in level analyses leads to broader coverage relative to growth analyses, particularly in earlier years. Second, datasets based on the best proxies are available for fewer countries. Third, coverage is smaller for $k = 3$ than for $k = -1$, since the condition $k < 2/(\mu - 1)$ always holds for $\mu > 1$ when $k < 0$, but fails when $k > 0$ and μ is sufficiently large. In particular, for $\mu > 1$ and $k = 3$, this condition is equivalent to $\mu < (2+k)/k \approx 1.67$. Finally, coverage

generally increases over time across all datasets.

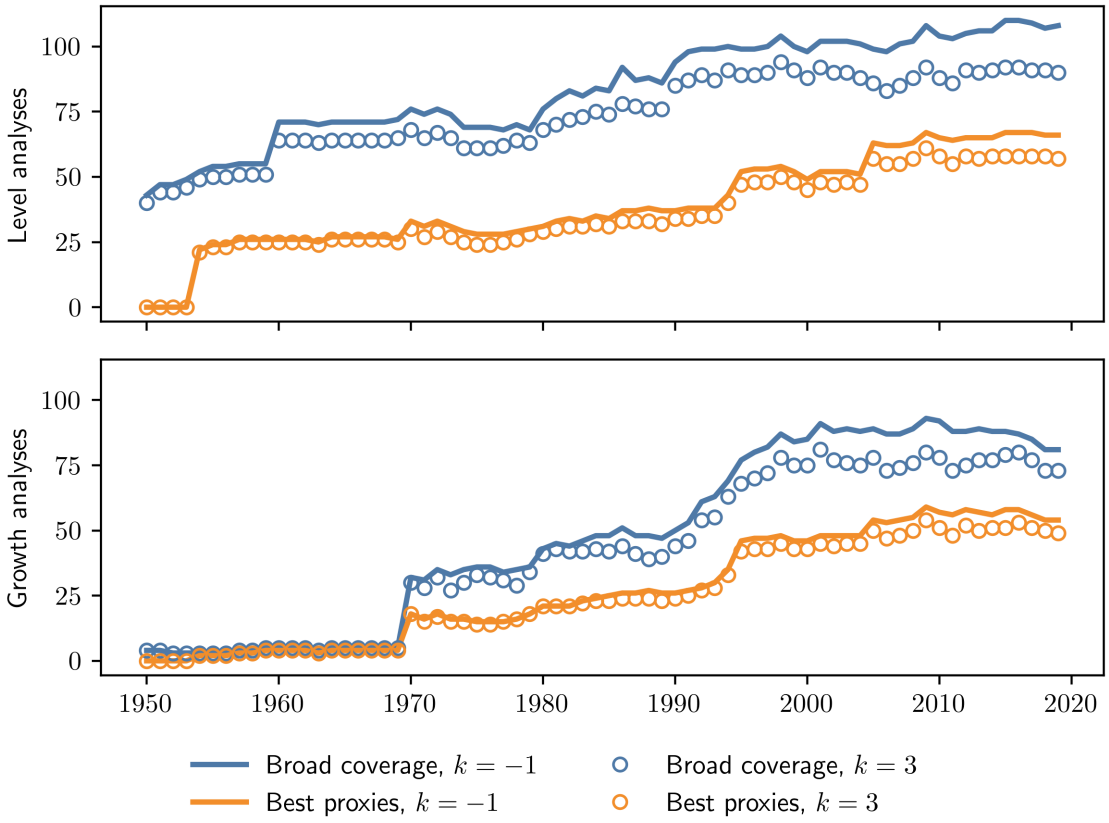


Figure 4: Number of non-oil countries with a valid TFP decomposition, 1950–2019.

In the subsequent sections, we leverage these datasets to revisit key aspects of economic growth through the lens of the TFP decomposition. As a first exploration of the datasets, Appendices F and G report scatter plots for levels and growth rates in selected years.

5 Growth at the frontier: US evidence

This section reexamines growth at the world income frontier. Following Jones (2016), we use the United States as a proxy for the frontier. The analysis relies on the growth dataset constructed using the most accurate available proxies, which provides coverage for the US starting in 1954.

The section proceeds in three steps. Section 5.1 validates the model by showing that our misallocation estimates are in line with those reported by Edmond et al. (2023), who employ a more flexible oligopoly framework. The remainder of the section is devoted to our empirical analysis. Section 5.2 reports estimates of average markup and misallocation, while Section 5.3 follows Jones (2016) in conducting a growth-accounting exercise for the US.

Taken together, these analyses assess the contribution of misallocation to both long-run productivity trends and short-run growth dynamics at the frontier.

5.1 Comparison with Edmond et al. (2023)

Edmond et al. (2023) estimate misallocation using an oligopoly model calibrated to the US economy. Rather than committing to a single value for the average markup μ, their calibration considers alternative values of $\mu \in \{1.05, 1.15, 1.25, 1.35\}$. Table 2 compares the value-added aggregate productivity losses reported in their Table 6 with our own estimates for $k = -1$ and $k = 3$, computed as $-(\overline{A}\Omega - \overline{A})/\overline{A} = 1 - \Omega$.

Table 2: Aggregate productivity losses, %

		Average markup μ			
		1.05	1.15	1.25	1.35
Edmond et al. (2023) 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

Notes: All entries report aggregate productivity losses (in percent). The first block reproduces the estimates of Edmond et al. (2023), derived from their oligopoly model; the second block reports our estimates for alternative values of k, computed as $1 - \Omega$.

As shown in Table 2, our estimates closely align with those of Edmond et al. (2023) when $k = -1$, particularly for $\mu = 1.15$, where the implied productivity losses are nearly identical. This correspondence is especially noteworthy given that their model is primarily calibrated using recent data—a period during which our estimates of μ remain close to 1.15 (Figure 5, first panel). For example, when targeting productivity dispersion—a key determinant of misallocation—their calibration focuses on concentration measures in 2012, a year for which we estimate $\mu = 1.18$.

This evidence provides an external validation for our framework, given that Edmond et al. (2023) employ a more general oligopoly model calibrated with firm-level data. Moreover, it suggests that $k = -1$ is a reasonable assumption for the US. Nonetheless, since $k = -1$ represents a lower bound in our model, we also report results for the upper bound $k = 3$ throughout the empirical analysis that follows.

5.2 Average markup and allocative efficiency estimates

Figure 5 plots our estimates of the cost-weighted average markup μ and the associated allocative efficiency Ω for the US, under $k = -1$ and $k = 3$. Several features of these series are worth highlighting.

First, reflecting the long-run decline in the labor income share LS, the average markup $\mu = (1 - \alpha)/LS$ exhibits a clear upward trend over time, albeit with distinct phases. In the early postwar decades, markups fluctuate around 1.10. They rise noticeably in the mid-1970s,

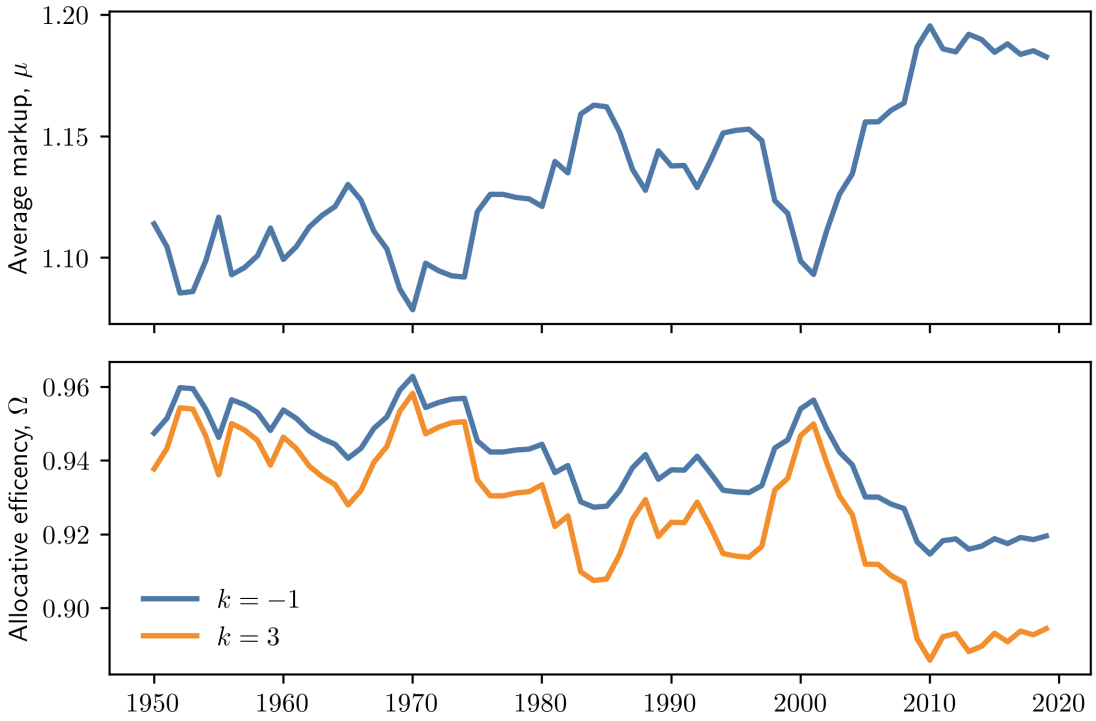


Figure 5: Cost-weighted average markup and allocative efficiency in the US, 1950–2019.

remaining around 1.15 until the mid-1990s, before declining sharply in the late 1990s. Markups rise markedly in the 2000s and stabilize around 1.18–1.20 in the 2010s. Overall, this pattern is consistent with a growing body of evidence documenting a increase in market power among US firms (e.g., De Loecker et al., 2020).[19]

Second, since allocative efficiency Ω is a decreasing function of μ, its evolution closely mirrors movements in average markups. Under the $k = -1$ specification, allocative efficiency declines gradually from approximately 0.95 in the 1950s and 1960s to around 0.92 after 2000, with a particularly pronounced deterioration during the 2000s. This pattern echoes the findings of Baqaee and Farhi (2020), who document an increase in the distance from the efficient allocation in the US between 1997 and 2015, driven by rising and increasingly dispersed markups.

Finally, consistent with Figure 1, allocative efficiency is higher and more stable for lower values of k. Quantitative differences across specifications remain relatively modest, especially in periods of low markups—such as the early part of the sample—when the mapping from μ to Ω is less sensitive to the choice of k.

Taken together, these results highlight a gradual deterioration in allocative efficiency over time, with important medium-run fluctuations that play a central role in the growth-accounting analysis that follows.

[19] Our estimated levels are lower and grow more gradually than those reported by De Loecker et al. (2020). These differences are consistent with the fact that their sample includes only listed firms, whose markups tend to be higher and to rise more over time (Díez et al., 2021).

5.3 Growth accounting

To understand the sources of economic growth, one possibility is to work directly with the aggregate production function (9). However, this formulation does not deliver a clean separation of growth components, since capital accumulation is itself partly induced by productivity growth—as in a standard Solow model. A better alternative, adopted by Jones (2016), is to use

$$\frac{Y_t}{E_t} = \left(\frac{K_t}{Y_t}\right)^{\frac{\alpha}{1-\alpha}} \left(\frac{L_t}{E_t}\right) \left(\overline{A}_t \Omega_t\right)^{\frac{1}{1-\alpha}}, \tag{16}$$

which follows from Equation (9) by introducing time subscripts, dividing through by Y_t^{α}, and rearranging terms. In this formulation, growth in output per hour Y_t/E_t arises through three distinct channels: (i) changes in the capital-output ratio, captured by $(K_t/Y_t)^{\frac{\alpha}{1-\alpha}}$; (ii) growth in human capital per hour, L_t/E_t, which in our case corresponds to the PWT human capital index; and (iii) growth in labor-augmenting TFP, $\left(\overline{A}_t \Omega_t\right)^{\frac{1}{1-\alpha}}$.

As emphasized by Jones (2016), this representation separates the contributions of productivity and capital in a way that the original production function (9) does not. By working with the capital-output ratio rather than the capital stock, it mitigates the mechanical comovement between capital accumulation and productivity.

The first four columns of Table 3 closely replicate the standard growth-accounting results reported in Table 3 of Jones (2016), based on Equation (16). Over the full sample, output growth is primarily driven by labor-augmenting TFP, with additional contributions from human capital accumulation and only a limited role for changes in the capital-output ratio. Looking across subperiods, growth is exceptionally strong between 1954 and 1973 amid outstanding productivity growth, possibly reflecting the long-run effects of the "great inventions" of the late nineteenth and early twentieth centuries (Gordon, 2000). Driven by TFP dynamics, growth slows markedly between 1973 and 1995, but accelerates again between 1995 and 2007, coinciding with the dot-com boom and the diffusion of information technology (Jones, 2016). Finally, TFP and output growth remain subdued after 2007. According to Fernald (2015), this productivity slowdown began before the Global Financial Crisis and reflects a retreat from the exceptional, but temporary, IT-fueled pace of the late 1990s and early 2000s. Similar conclusions are reached by Byrne et al. (2013).

These results underscore the central role of TFP in explaining economic growth in both the short and the long run. Our framework allows us to go beyond this well-established conclusion by decomposing labor-augmenting TFP into technology and allocative efficiency components. The last four columns of Table 3 report this decomposition for $k = -1$ and $k = 3$. In light of the evidence in Table 2, we focus on the results for $k = -1$; nevertheless, our conclusions are not materially affected when using $k = 3$.

The decomposition reveals that long-run TFP growth is overwhelmingly driven by technological progress. Between 1954 and 2019, labor-augmenting TFP grows at 1.1 percent per year,

Table 3: Growth accounting for the US

Period	GDP/hour	GDP/hour decomp.			Labor-aug. TFP decomp.			
					$k=-1$		$k=3$	
	$\frac{Y}{E}$	$\left(\frac{K}{Y}\right)^{\frac{\alpha}{1-\alpha}}$	$\frac{L}{E}$	$\left(\overline{A}\Omega\right)^{\frac{1}{1-\alpha}}$	$\overline{A}^{\frac{1}{1-\alpha}}$	$\Omega^{\frac{1}{1-\alpha}}$	$\overline{A}^{\frac{1}{1-\alpha}}$	$\Omega^{\frac{1}{1-\alpha}}$
1954–2019	1.8	0.2	0.5	1.1	1.2	-0.1	1.2	-0.1
1954–1973	2.6	-0.0	1.0	1.7	1.7	0.0	1.7	0.0
1973–1995	1.3	0.3	0.5	0.5	0.7	-0.2	0.8	-0.3
1995–2000	2.5	0.2	0.3	2.0	1.3	0.7	1.0	1.0
2000–2007	2.1	0.4	0.3	1.4	1.9	-0.6	2.2	-0.8
2007–2019	1.0	0.2	0.2	0.6	0.8	-0.1	0.8	-0.2

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

nearly matching the 1.2 percent annual growth rate of the technology component. Changes in allocative efficiency play a minor role over this horizon, exerting a small drag on growth, in line with the gradual deterioration in Ω documented in Figure 5.

Misallocation, however, becomes substantially more relevant over shorter horizons. During the IT boom, TFP grows rapidly in both 1995–2000 and 2000–2007, but for fundamentally different reasons. From 1995 to 2000, strong TFP growth reflects sizable contributions from both components—technological progress and improvements in allocative efficiency. By contrast, in 2000–2007, TFP growth is driven entirely by technology, as allocative efficiency deteriorates throughout the period.

This pattern is consistent with sector-level evidence. Fernald (2015) documents large productivity gains in IT-producing industries during 1995–2000, while IT-intensive sectors experienced only modest improvements. Similarly, Gordon (2000) shows that productivity gains in 1995–1999 were primarily concentrated in computer and semiconductor manufacturing. Given that IT-producing sectors accounted for only 4.9 percent of private business value added between 1988 and 2011 (Fernald, 2015), such sector-specific gains would not mechanically translate into large aggregate TFP growth.

While the magnitude of aggregate TFP growth in the late 1990s may therefore appear difficult to reconcile with this sectoral evidence, our decomposition helps address this tension. The pace of technological progress in 1995–2000 is only slightly above its 1954–2019 average, whereas the unusually high TFP growth during this period is instead driven by rapid improvements in allocative efficiency. This pattern suggests that the late-1990s TFP surge reflects not only innovation within a narrow set of IT-producing sectors, but also—and primarily—a more efficient allocation of resources, consistent with a reduction in the cost of technology adoption.

We find sizable technology improvements only after 2000, consistent with the evidence in

Fernald (2015) that TFP in *large* IT-intensive industries surged in the early twenty-first century. Remarkably, technological progress is even faster in 2000–2007 than in 1954–1973, despite higher aggregate TFP growth in the earlier period. Nevertheless, the simultaneous decline in allocative efficiency during 2000–2007 suggests that these IT-driven technological gains disproportionately accrued to a subset of firms.

This interpretation aligns with the evidence in Autor et al. (2020), who document rising firm concentration in sales and employment around 2000. A plausible mechanism underlying this pattern is the growing importance of intangible inputs, such as information technology and software. When firms differ in their ability to deploy such intangibles efficiently—as supported by micro-level evidence in De Ridder (2024)—these inputs confer a significant competitive advantage to some firms, discouraging entry and expansion by others. De Ridder (2024) argue that this mechanism can jointly account for rising concentration, the temporary acceleration of productivity growth, and its subsequent slowdown. Related insights are provided by Aghion et al. (2023).

Altogether, these results suggest that misallocation plays a limited role in shaping long-run TFP growth, but can have sizable effects over shorter horizons. The IT-boom period illustrates how short-run fluctuations in allocative efficiency were quantitatively important. More broadly, misallocation appears to account for a substantial share of the variation in TFP growth across subperiods. By contrast, the growth rate of the technology component is relatively stable over time, hovering around 1 percent per year, except for two episodes—1954–1973 and 2000–2007—when technological progress accelerates to roughly 2 percent per year. This pattern is consistent with the interpretation advanced by Fernald (2015), who argues that these two episodes of rapid productivity growth—rather than the subsequent slowdowns—are the true unusual events. The slowdowns essentially reflect a return to a more normal growth pace.

6 Cross-country income differences: a development accounting exercise

Section 5.3 relies on a growth accounting exercise to identify the proximate determinants of economic growth in the US, taken as a proxy for the global income frontier. We now turn to development accounting, applying the same logic to explain cross-country income differences at a given point in time. This exercise is conducted for each year between 1950 and 2019. As in Section 5.3, we rely on production function (16) rather than (9), since the former allows for a clearer separation between the contributions of productivity and capital.[20]

As Jones (2016) points out, the capital-output ratio is remarkably stable across countries. Human capital varies more, as poorer countries tend to exhibit lower levels of educational at-

[20]In the context of this section, the subscript t in (16) should be interpreted as indexing countries rather than time.

tainment, but this variation remains modest relative to the large observed cross-country income differences. These facts imply that cross-country income disparities primarily reflect productivity differences, leading to the standard conclusion that income differences are predominantly attributed to TFP (Caselli, 2005, 2016; Jones, 2016). This raises the question that motivates this section: does misallocation offer additional insights into cross-country income differences?

To answer this question, we employ standard measures of success from the development-accounting literature, which are based on a variance decomposition of income across countries. We compute these measures under two alternative specifications: (i) treating TFP $\overline{A}\Omega$ as the residual, as in conventional development accounting; and (ii) treating only the technology component, $\overline{A}$, as the residual, consistent with the TFP decomposition implied by our model. Comparing these two specifications allows us to quantify the role of misallocation in explaining cross-country income differences.

6.1 Methodology

To assess the extent to which misallocation helps explain cross-country income differences, we explore the variance decomposition implied by Equation (16):

$$\mathrm{Var}(y) = \mathrm{Var}(e) + \mathrm{Var}(u) + 2\mathrm{Cov}(e,u), \tag{17}$$

where $\mathrm{Var}(\cdot)$ and $\mathrm{Cov}(\cdot)$ denote the sample variance and covariance across countries in a given year. Here, $y \equiv \ln(Y/E)$ is the natural log income per unit of raw employment, e represents the explained component of y—the "observable" determinants in the production function—and u denotes the unexplained component, i.e., the residual.

We consider two alternative specifications for the explained and unexplained components. First, as in standard development-accounting exercises, capital and labor are treated as observables and TFP as the residual. In this case,

$$e \equiv \ln\left[(K/Y)^{\frac{\alpha}{1-\alpha}}(L/E)\right] \quad \text{and} \quad u \equiv \ln\left[\left(\overline{A}\Omega\right)^{\frac{1}{1-\alpha}}\right].$$

Second, consistent with our oligopoly model, we include allocative efficiency Ω in the explained component, so that the residual captures only the technology component of TFP, $\overline{A}$. Accordingly,

$$e \equiv \ln\left[(K/Y)^{\frac{\alpha}{1-\alpha}}(L/E)\,\Omega^{\frac{1}{1-\alpha}}\right] \quad \text{and} \quad u \equiv \ln\left[\overline{A}^{\frac{1}{1-\alpha}}\right].$$

For each specification, we use (17) to compute the share of cross-country income variance accounted for by the observables. We report two standard measures of success from the

development-accounting literature (Klenow and Rodríguez-Clare, 1997; Caselli, 2005):

$$success_1 \equiv \frac{\mathrm{Var}(e)}{\mathrm{Var}(y)} \tag{18}$$

$$success_2 \equiv \frac{\mathrm{Var}(e) + \mathrm{Cov}(e,u)}{\mathrm{Var}(y)} = \frac{\mathrm{Cov}(e,y)}{\mathrm{Var}(y)}. \tag{19}$$

Both measures incorporate $\mathrm{Var}(e)$ in the numerator, but they differ in their treatment of the covariance term in (17). The first measure attributes the entire covariance term to the unexplained component, whereas the second assigns half of it to the explained component. The latter can also be interpreted as the OLS slope coefficient from regressing e on y. As emphasized by Klenow and Rodríguez-Clare (1997), this measure amounts to asking: when income per unit of raw employment is 1 percent higher in one country, how much higher do we expect the explained component to be?

High values of these measures indicate that the observables successfully account for cross-country income differences, while low values imply that the residual plays a dominant role. By computing these metrics both with and without allocative efficiency Ω included in e, we assess whether misallocation enhances our understanding of cross-country income differences.

6.2 Results

Figure 6 reports the success measures for the period 1950–2019. The top row uses the level dataset with broad coverage, while the bottom row relies on the level dataset with best proxies. Panels in the left column report the first success measure (Equation (18)), whereas those in the right column report the second measure (Equation (19)). Within each panel, we compare three specifications for the explained component e. The dashed blue line corresponds to the standard case in which only inputs enter e. The solid orange and green lines additionally include allocative efficiency Ω in e, with Ω estimated assuming a Pareto shape parameter of $k = -1$ and $k = 3$, respectively.

For each dataset and year, we impose a common country sample, retaining only non-oil countries for which a valid TFP decomposition exists under $k = 3$. Consequently, only countries with $1 < \mu < 1.67$ are included in the analysis.[21] This restriction ensures comparability of the series across values of k within each dataset and year.[22] Comparisons across datasets or over time should be interpreted with caution, since sample coverage varies both over time within a given dataset and across datasets for a given year.

As shown in Figure 6, both success measures increase when misallocation is included in the explained component e across all evaluated years. This pattern is even more evident in Figure

[21] The necessary condition $k < \frac{2}{\mu-1}$ is always satisfied for $k < 0$, but fails for $k > 0$ when μ is sufficiently large. In particular, for $\mu > 1$ and $k = 3$, this condition is equivalent to $\mu < \frac{2+k}{k} \approx 1.67$.

[22] Our conclusions are generally robust to allowing the sample to vary with k, using all non-oil countries available in each case.

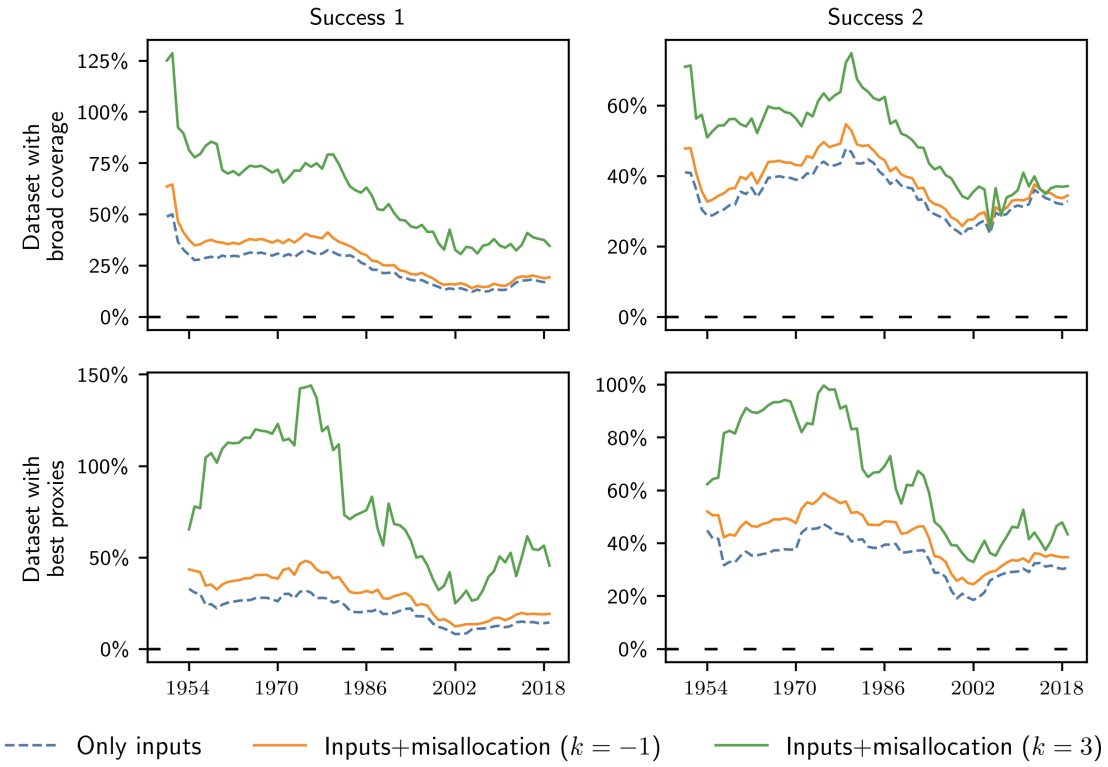


Figure 6: Measures of success, common sample of non-oil countries for each dataset and year, 1950–2019.



Figure 7: Success gains due to misallocation, common sample of non-oil countries for each dataset and year, 1950–2019.

7, which plots the difference between the measures computed with and without misallocation included in e. The gains are more pronounced for the first measure, indicating that while misallocation increases both the variance of the explained component (the numerator of Equation (18)) and its covariance with income levels (the numerator of Equation (19)), the former rises more sharply.

The gains are generally smaller under $k = -1$, typically below 10 percentage points, and larger under $k = 3$. This difference follows directly from the model's structure. As shown in Figure 1, allocative efficiency Ω is a decreasing function of the cost-weighted average markup μ, and higher values of the Pareto shape parameter k make Ω more sensitive to differences in μ. These properties imply that a higher value of k (i) increases the cross-country dispersion of allocative efficiency Ω and (ii), since the average markup μ is negatively related to the capital-labor composite $K^{\alpha}L^{1-\alpha}$, strengthens the positive cross-country association between Ω and the remaining observables. Consequently, when allocative efficiency is included in the explained component e, both the variance of Ω and its covariance with the capital-labor composite contribute to raising $\mathrm{Var}(e)$, and hence the first measure of success. The same mechanisms help account for the increase in the second measure under $k = 3$. Since markups are negatively correlated with income (Figures F.13, F.14, F.15, and F.16 in Appendix F), increases in k tend to amplify the positive cross-country association between allocative efficiency and income (Figures F.9, F.10, F.11, and F.12 in Appendix F), thereby raising the numerator of the second measure, $\mathrm{Cov}(e, y)$.

The gains appear stronger when using the dataset with best proxies, roughly doubling relative to the broad-coverage dataset. However, this pattern largely reflects differences in country coverage rather than a genuine effect of using alternative measures for the underlying variables. When we restrict attention to a common country sample across datasets, these higher gains mostly disappear (Figures H.1 and H.2 in Appendix H). Similarly, the observed evolution of the success measures over time is sensitive to changes in country coverage. Fixing a common sample over time typically alters the time-series patterns and can be highly dependent on the choice of sample window (see Figures H.3 and H.4 in Appendix H, which report results under a time-invariant country sample from 1960 onwards).

Overall, these results indicate that market-power-driven misallocation plays a meaningful role in explaining cross-country income differences, particularly under higher values of k. Nevertheless, as shown in Figure 6, the success measures remain relatively low—typically below 50 percent in recent decades, when country coverage is broader. As noted by Caselli (2005), a common summary of the development-accounting literature is that differences in TFP account for at least 50 percent of differences in per capita income. Our results suggest that while misallocation enhances our understanding of income differences, it does not fundamentally overturn this central conclusion: the majority of cross-country income variation remains essentially unexplained.

7 Economic convergence

Section 6 examined the role of misallocation in explaining cross-country income differences at each point in time. In this section, we maintain a global perspective but shift the focus to dynamics, assessing whether the data provide evidence of economic convergence.

Jones (2016) evaluates income convergence by relating average economic growth over 1960–2011 to GDP per capita in 1960 (see Figure 26 of that paper). Since these variables do not exhibit a negative relationship, he concludes that "a simplistic view of convergence does not hold for the world as a whole." In Figure 8, we conduct a similar exercise for the periods 1975–1995, 1975–2005, 1975–2015, 1985–2005, 1985–2015, and 1995–2015. In each panel, the y-axis shows the annual growth rate of GDP per capita for the corresponding period, drawn from the growth dataset with broad coverage, while the x-axis reports income levels in the initial year of the period, taken from the level dataset with broad coverage. Each exercise includes all countries with a valid TFP decomposition under $k = -1$ in both the level and growth datasets for the corresponding period. We distinguish between oil-producing and non-oil-producing countries in the plots.[23] For each period, we estimate cross-country OLS regressions of average annual income-per-capita growth (in percent) on initial log income-per-capita levels. A negative slope is therefore indicative of convergence: for example, a coefficient of -0.5 implies that a country with an initial income level 10 percent higher grows about 0.05 percentage points more slowly per year on average. The estimated slope coefficients and their statistical significance—for non-oil countries and for the full set of countries—are reported within each panel. Consistent with Jones (2016), we find little evidence in favor of income convergence. A negative relationship between growth and initial income emerges only for the most recent period, 1995–2015.[24]

Section 6.2 showed that TFP plays a central role in accounting for cross-country income differences, suggesting that the lack of income convergence may reflect limited convergence in productivity. We investigate this hypothesis in Figures 9, 10, and 11, which report convergence assessments for TFP $\overline{A}\Omega$, the technology frontier $\overline{A}$, and allocative efficiency Ω, respectively. These exercises mirror Figure 8; in particular, they rely on the broad-coverage level and growth datasets, including all countries with a valid TFP decomposition under $k = -1$ in both datasets for the corresponding period. As in the income case, the convergence hypothesis finds little support for TFP and for its technology component except possibly for the 1995–2015 period, when poorer countries experienced faster growth amid rapid technological progress. In contrast, the evidence is consistent with convergence in allocative efficiency. We find a negative relationship between allocative-efficiency growth and its initial level across all evaluated periods for non-oil countries. A similar pattern holds for the full set of countries, with the exception of the 1985–2005 period.

[23] An oil-producing country is defined here as one in which oil rents exceed 10 percent of GDP in *any* year within the corresponding period.

[24] We follow Jones (2016) and evaluate income per capita in this exercise. Using income per unit of raw employment instead yields very similar results and does not affect the qualitative conclusions.

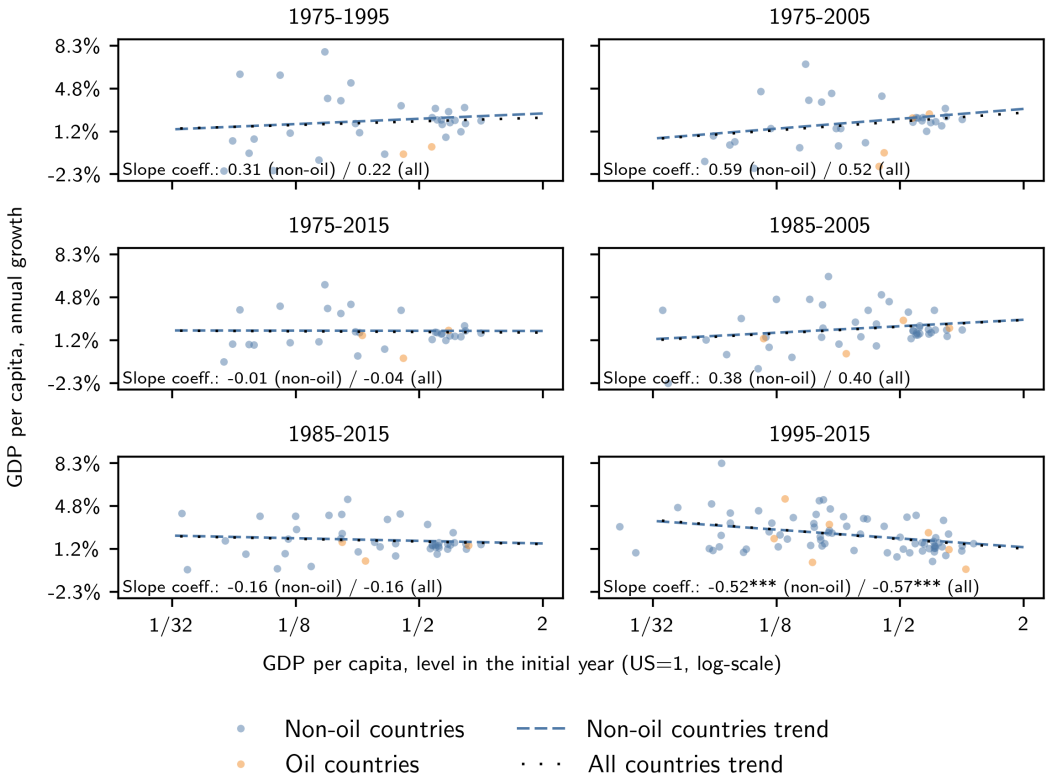


Figure 8: Assessing convergence in income per capita – datasets with broad coverage ($k = -1$). *Note*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

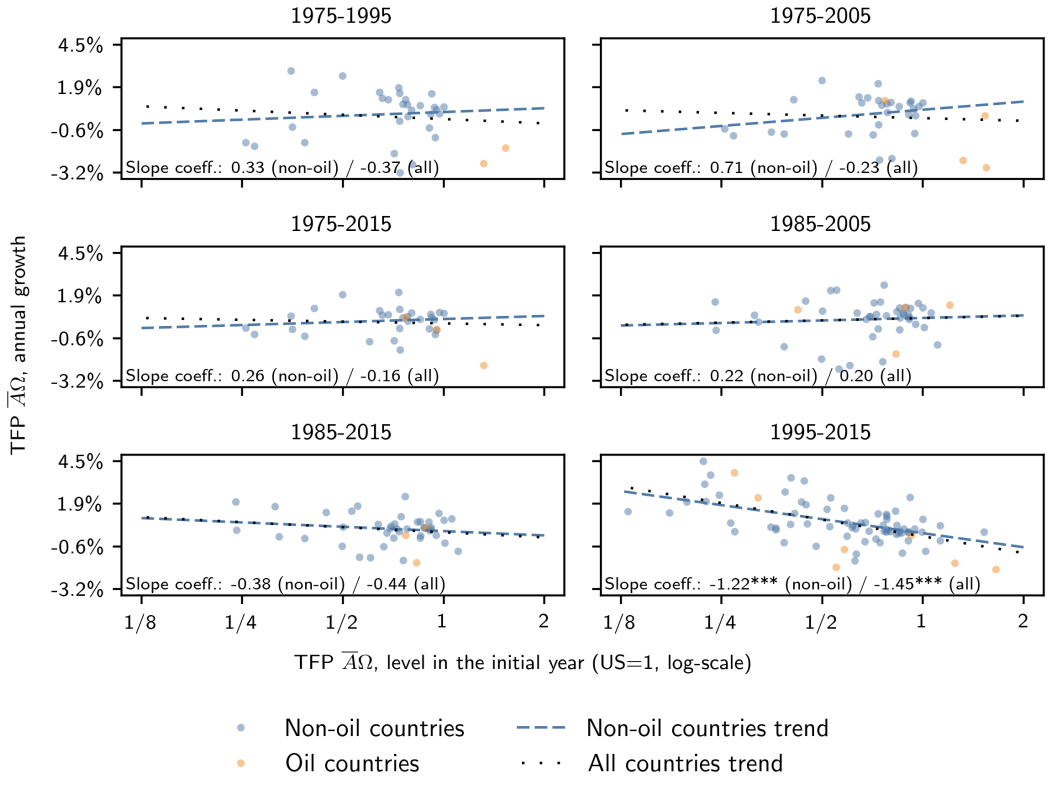


Figure 9: Assessing convergence in TFP $\overline{A}\Omega$ – datasets with broad coverage ($k = -1$). *Note*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

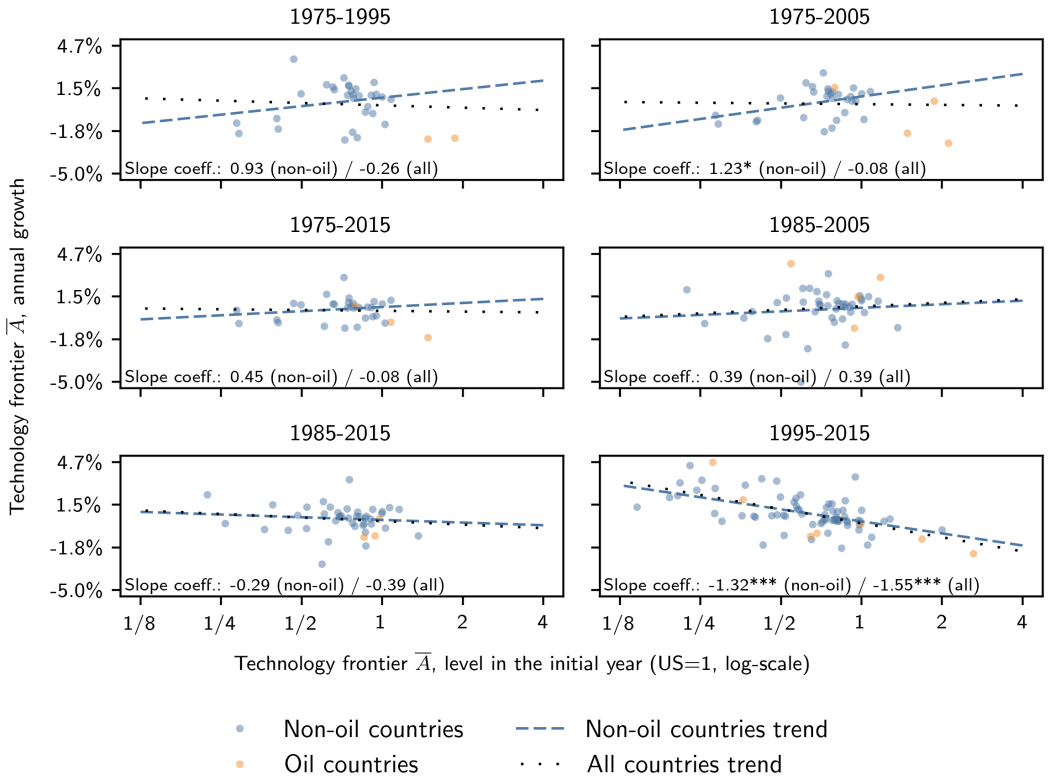


Figure 10: Assessing convergence in the technology frontier $\overline{A}$ – datasets with broad coverage ($k = -1$). *Note*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

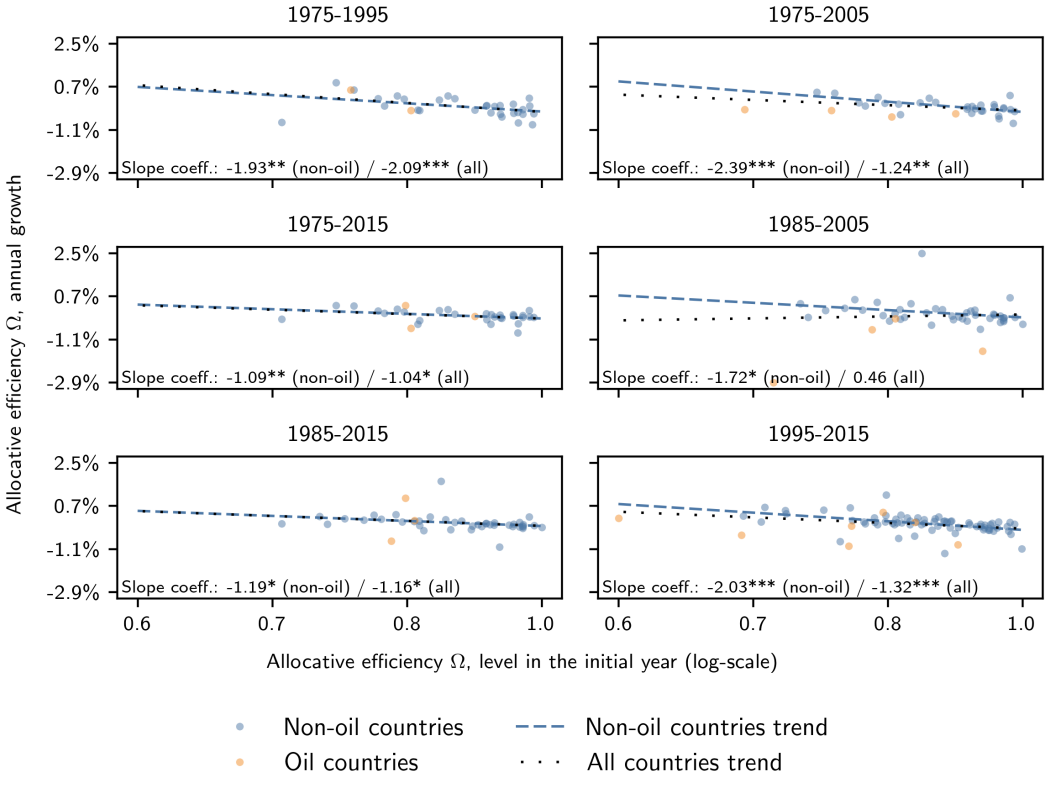


Figure 11: Assessing convergence in allocative efficiency Ω – datasets with broad coverage ($k = -1$). *Note*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A more comprehensive convergence assessment is presented in Figure 12, which reports the same slope estimates shown in the scatter plots, now computed for each 20-year window starting in 1970, together with their 90 percent confidence intervals. Overall, the evidence reinforces our main findings. Convergence in income, TFP $\overline{A}\Omega$, and the technology frontier $\overline{A}$ is largely absent, except possibly in more recent decades. By contrast, the estimates provide evidence of convergence in allocative efficiency Ω among non-oil countries, as a negative relationship emerges in most periods. Moreover, the slope coefficient for non-oil countries is economically sizable, typically around -2, implying that a 10 percent higher initial level of allocative efficiency is associated with allocative-efficiency growth about 0.2 percentage points lower per year on average. For the full set of countries, the estimated coefficients are generally negative but not statistically significant.

These conclusions are generally robust across datasets and values of k. The absence of convergence in income, TFP, and the technology frontier holds both for the broad-coverage datasets under $k = 3$ (Figure 13) and for the datasets with best proxies under either $k = -1$ or $k = 3$ (Figures 14 and 15, respectively). Evidence of convergence in allocative efficiency is also broadly present, particularly for non-oil countries, albeit with an important qualification. Specifically, under $k = 3$, the estimated slopes tend to lose statistical significance in more recent decades, especially when using the datasets with best proxies.[25]

In sum, the empirical evidence provides little support for convergence in income or in aggregate TFP, with the latter reflecting the absence of convergence in its technology component. By contrast, we find suggestive evidence of convergence in allocative efficiency, although this pattern is not uniform across all periods, country samples, datasets, and values of k. Taken together, these findings suggest that barriers to international technology diffusion persist over long periods, limiting convergence in productivity and income across countries. At the same time, within-country forces—such as competition or market integration—may gradually reduce market-power-driven misallocation. In this sense, allocative efficiency appears more amenable to convergence than the technology frontier, whose evolution remains closely tied to long-lasting country-specific factors, including institutions.

8 Robustness checks

Throughout the paper, we have already examined the sensitivity of our results to key modeling choices. In particular, we considered alternative calibrations for the Pareto shape parameter k and applied the methodology to different datasets. In this section, we implement two additional robustness checks.

[25]This pattern is not driven by differences in country coverage across values of k. The evidence continues to support convergence in allocative efficiency among non-oil countries under $k = -1$ when the sample is restricted to countries with a valid TFP decomposition under $k = 3$. In fact, under this restricted sample, we find consistent evidence of convergence in Ω even when using the broad-coverage datasets for the full set of countries under $k = -1$, in contrast to the results reported in Figure 12.

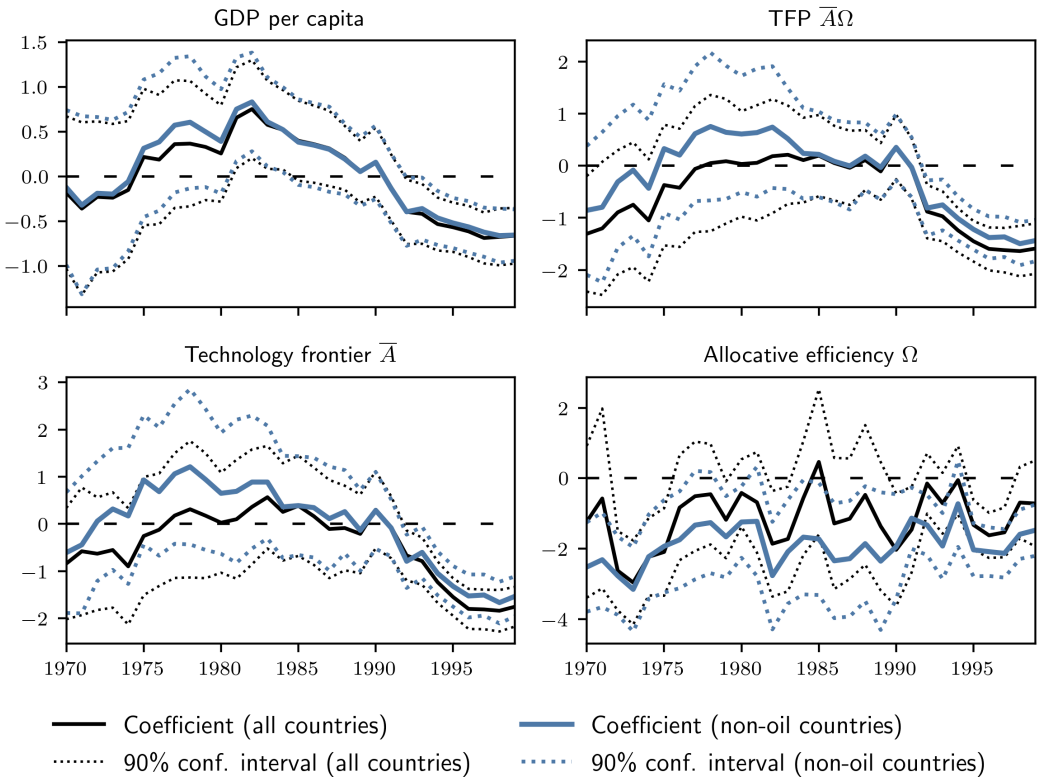


Figure 12: Slope coefficient from regressing 20-year-ahead annual growth (in percent) on initial log levels, 1970–1999 – datasets with broad coverage ($k = -1$).

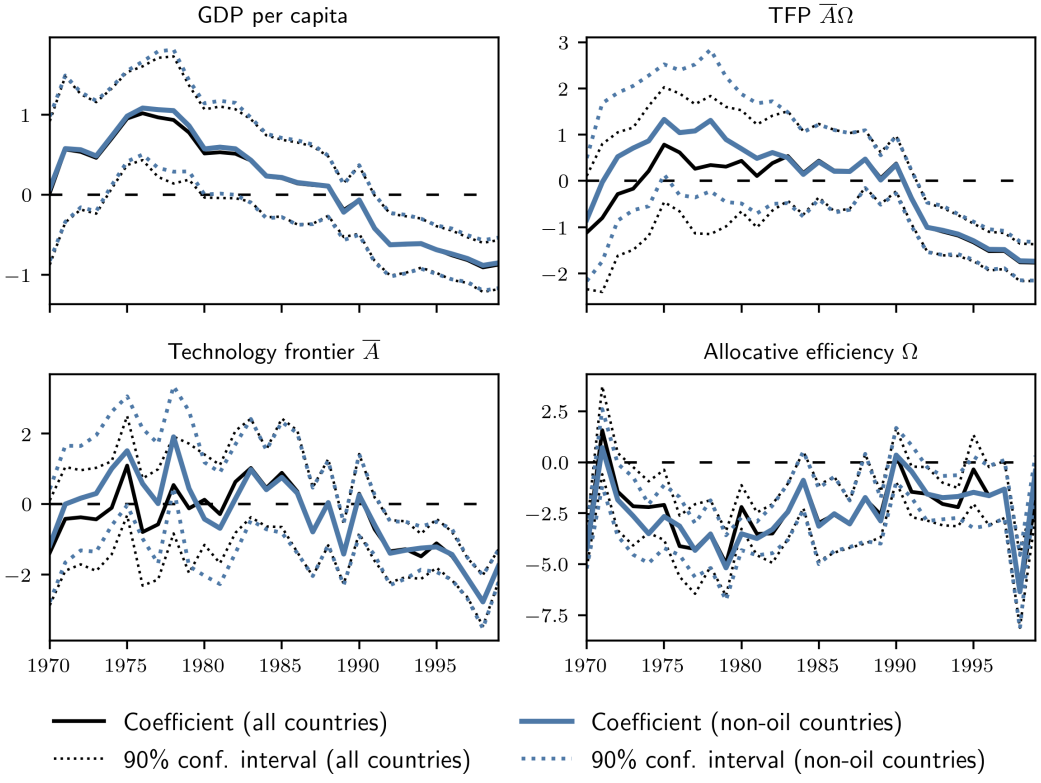


Figure 13: Slope coefficient from regressing 20-year-ahead annual growth (in percent) on initial log levels, 1970–1999 – datasets with broad coverage ($k = 3$).

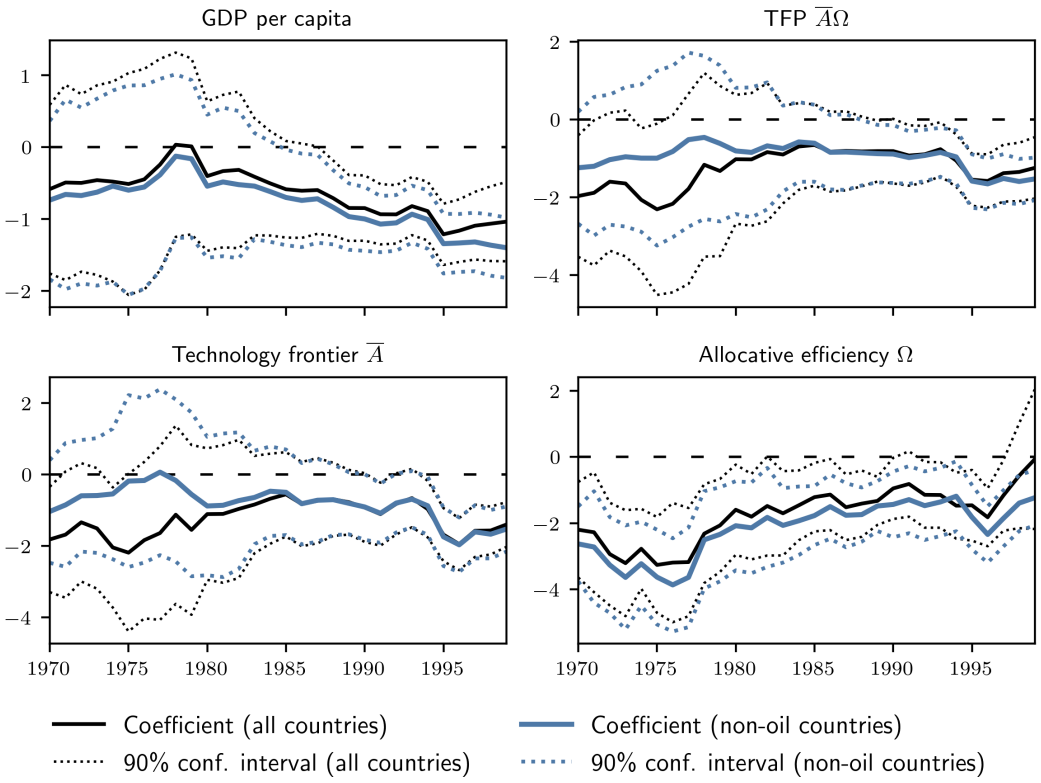


Figure 14: Slope coefficient from regressing 20-year-ahead annual growth (in percent) on initial log levels, 1970–1999 – datasets with best proxies ($k = -1$).

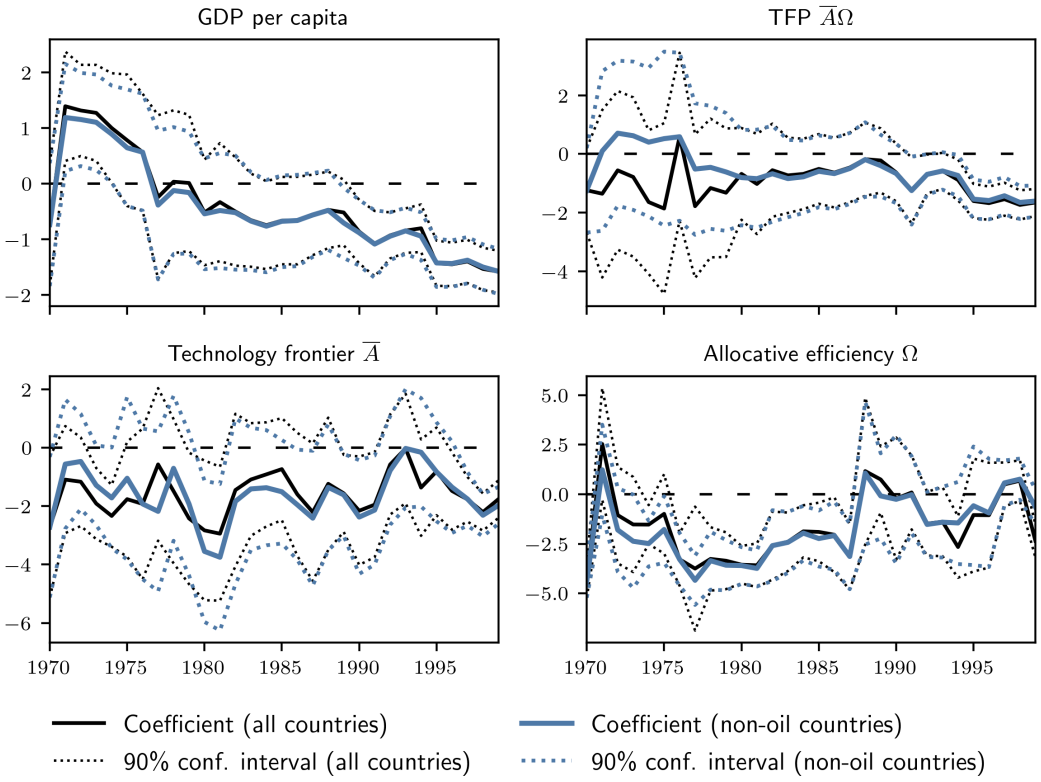


Figure 15: Slope coefficient from regressing 20-year-ahead annual growth (in percent) on initial log levels, 1970–1999 – datasets with best proxies ($k = 3$).

A natural concern relates to parameter choices. Having already explored economically plausible bounds for k, we focus here on the capital elasticity α. Although our baseline choice $\alpha = 0.30$ is grounded in empirical evidence, some degree of uncertainty inevitably remains. Accordingly, in our first robustness exercise, we show that the main empirical results are essentially unaffected by alternative plausible calibrations of the capital elasticity.

Second, our cross-country analyses have focused primarily on non-oil countries. An important question is whether our conclusions extend to the full sample of countries. While the convergence analysis in Section 7 already addressed this issue, the development-accounting exercise of Section 6 was conducted exclusively for non-oil countries. Our second robustness check repeats that exercise for the full sample, yielding qualitatively similar conclusions.

Alternative calibrations of the capital elasticity. Instead of relying on our baseline choice $\alpha = 0.30$, we consider two alternative values, $\alpha = 1/3$ and $\alpha = 0.27$. As discussed in Section 3.3, we constructed plausible bounds for the capital elasticity in the global economy, in the spirit of the US bounds proposed by Vollrath (2024). The resulting global and US bounds, reported in Appendix D, indicate that these alternative values of α correspond to relatively extreme cases within the plausible parameter space (see Figure D.1).[26]

The results under these alternative values of the capital elasticity are reported in Appendix I. Tables I.1 and I.2 replicate the US growth-accounting exercise presented in Table 3, using $\alpha = 1/3$ and $\alpha = 0.27$, respectively. The results are virtually identical.[27]

Figures I.1 and I.2 report, respectively, the development-accounting success measures and the corresponding success gains due to misallocation for $\alpha = 1/3$. Figures I.3 and I.4 present the analogous results for $\alpha = 0.27$. As in Figures 6 and 7, both success measures increase when misallocation is included in the explained component e. The gains are more pronounced for the first measure and, for both measures, under $k = 3$ than under $k = -1$. Nevertheless, as in the baseline case, the success measures typically remain below 50 percent in recent decades.

Finally, Figures I.5, I.6, I.7, and I.8 reassess economic convergence under $\alpha = 1/3$ for each value of k and each dataset. The corresponding results for $\alpha = 0.27$ are shown in Figures I.9, I.10, I.11, and I.12. As in the baseline specification, we find little evidence of convergence in income, TFP, or the technology frontier. By contrast, the evidence remains consistent with convergence in allocative efficiency, particularly for non-oil countries and under $k = -1$.

[26] Although $\alpha = 1/3$ is among the most common assumptions in economics (Vollrath, 2024), it lies above the no-profits upper bound for the US over most of the sample period, falling below this bound only from the early 2000s onward. This calibration is adopted in Edmond et al. (2023) and is also widely used in development-accounting exercises (Caselli, 2005, 2016; Jones, 2016).

[27] Adopting a higher (lower) value of α shifts the cost-weighted average markup $\mu = (1 - \alpha)/LS$ downward (upward), and therefore shifts allocative efficiency Ω upward (downward). For instance, under $k = -1$, we estimate Ω equal to 0.92 for the baseline $\alpha = 0.30$, 0.94 for $\alpha = 1/3$, and 0.90 for $\alpha = 0.27$. However, these differences primarily reflect level shifts and thus do not materially affect the growth rates of Ω.

Non-oil countries versus all countries. Section 7 evaluated economic convergence both for non-oil countries and for the full sample of countries. Overall, the results are broadly similar, although the evidence of convergence in allocative efficiency is stronger when attention is restricted to non-oil countries. By contrast, the development-accounting exercise in Section 6 focused exclusively on non-oil economies.

To assess the implications of extending this analysis to the full sample, Appendix J repeats the development-accounting exercise including oil-producing countries. The corresponding success measures and success gains due to misallocation are reported in Figures J.1 and J.2, respectively. The qualitative conclusions remain unchanged: misallocation generates positive success gains—particularly for the first measure and under $k = 3$—yet the success measures typically remain below 50 percent in recent decades.

9 Model extension: beyond Cobb-Douglas production functions

Appendix C shows that our main theoretical results do not hinge on the Cobb-Douglas production. Instead, they hold for any well-behaved production function with an arbitrary number of inputs, provided that it satisfies (i) constant returns to scale and (ii) Hicks-neutral productivity. Under these conditions, the quantitative strategy for TFP decomposition described in Section 3 remains valid. Specifically, given the shape of the productivity distribution, decomposing TFP requires targeting only two moments: (i) aggregate TFP, $\overline{A}\Omega$, and (ii) the cost-weighted average markup, μ. The calibration proceeds exactly as before, following the algorithm described in Section 3.1. Importantly, the existence and uniqueness results established for the Cobb-Douglas case in Appendix B continue to hold in this more general setting. In short, the mapping from the average markup μ to allocative efficiency Ω—and hence the entire TFP decomposition—remains unchanged.

The only substantive difference introduced by this generalization concerns the empirical computation of the target moments. Aggregate TFP is still identified as a residual from the aggregate production function, although its numerical value depends on the chosen functional form. Likewise, the cost-weighted average markup μ continues to be pinned down by the ratio of the labor elasticity of output to the labor income share. However, while in the Cobb-Douglas case this elasticity is constant and equal to $1 - \alpha$, under more general technologies, it depends on relative factor quantities.[28]

[28]Naturally, computing these moments requires calibrating the production function. In Section 3.3, we discuss how this calibration can be performed using cost data in the Cobb-Douglas case, where $\alpha = \frac{Kr}{Kr+Lw}$. Appendix C.3 shows how a similar cost-based approach extends to the generalized production function.

10 Conclusion

This paper develops a tractable Cournot framework that shows how standard macroeconomic data can be used to discipline models of market power and misallocation. We implement this framework to decompose aggregate TFP into technology and allocative efficiency components for up to one hundred countries over the period 1950–2019, and use this decomposition to revisit key facts of economic growth.

At the world income frontier, proxied by the United States, we find that misallocation plays a limited role in shaping long-run TFP growth but can have sizable effects over shorter horizons. During the 2000–2007 period, for instance, the US experienced rapid technological progress alongside declining allocative efficiency, consistent with the view that the diffusion of information technology generated productivity gains, but these were concentrated among a subset of highly productive firms. More broadly, allocative efficiency accounts for a substantial share of the variation in TFP growth across subperiods, whereas the growth rate of the technology component is relatively stable over time, hovering around 1 percent per year, except for two notable episodes—1954–1973 and 2000–2007—during which technological progress accelerates to roughly 2 percent per year.

From a cross-country perspective, accounting for misallocation improves our understanding of income differences. Both success measures from the development-accounting exercise increase when allocative efficiency is included in the explained component. The gains are larger for the first measure—which is based on the variance of the explained component—and, for both measures, under higher values of the Pareto shape parameter k. Nevertheless, these improvements do not fundamentally overturn the central conclusion of the development-accounting literature: most cross-country income variation remains essentially unexplained.

Finally, we find little support for convergence in income or aggregate TFP. The absence of TFP convergence reflects a lack of convergence in the technology component, while we find suggestive—but not uniform—evidence of convergence in allocative efficiency. This pattern is consistent with the view that barriers to international technology diffusion persist over long periods, limiting convergence in productivity and income, whereas within-country forces—such as competition or market integration—may gradually reduce market-power-driven misallocation.

Appendices

A Model: proofs and derivations

Section 2 presents the model with a focus on its economic content, omitting formal proofs and derivations. The objective of this appendix is to provide those results. In what follows, let $g_a(A) \equiv \frac{g(A)}{1-G(\underline{A})}$ denote the density function of active firms—i.e., for $A \in [\underline{A}, \overline{A}]$—with cumulative distribution function G_a.

As a first step, we present a simple but useful mathematical result that serves as a building block for several key results of the model.

Proposition A.1 *Let h be a function of A such that (i) $\frac{\partial h(A)}{\partial \underline{A}} = 0$ and (ii) $\mathrm{E}_a\left(h(A)\right)$ is well defined. Then,*

$$\frac{\partial \mathrm{E}_a\left(h(A)\right)}{\partial \underline{A}} = g_a(\underline{A})\left[\mathrm{E}_a\left(h(A)\right) - h(\underline{A})\right].$$

Proof. Given a function h satisfying the two conditions in the statement,

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

where we use the fact that $\frac{\partial\left[\int_{\underline{A}}^{\overline{A}} h(A)g(A)dA\right]}{\partial \underline{A}} = -h(\underline{A})g(\underline{A})$, since h does not depend on $\underline{A}$. ■

A.1 Solving for the equilibrium allocation

Using Equation (6) for any active firms i and j, and noting that goods are homogeneous with equalized prices across firms within a sector, we can show that

$$A_i\left(\eta - s_i\right) = A_j\left(\eta - s_j\right), \tag{A.1}$$

where we drop the sectoral subscript since, as argued in Section 2, sectors are ex ante and ex post equivalent. Rewriting it as $\frac{s_i}{A_j} = \frac{\eta}{A_j} - \frac{1}{A_i}\left(\eta - s_j\right)$ and summing in j over all active firms—that is, all firms j with $A_j \geq \underline{A}$—we obtain:

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

Equation (A.2) implies that not all firms can be simultaneously active when N becomes sufficiently large: if $N_a = N \to +\infty$, then $s(\underline{\underline{A}}) < 0$, since $\underline{\underline{A}} < \overline{A}$. Formally, some firms must remain inactive if N exceeds the threshold $\check{N}$, which, from Equation (A.2), is given by

$$0 = \frac{1}{\check{N}\mathrm{E}(\underline{\underline{A}}/A)} + \eta\left(1 - \frac{1}{\mathrm{E}(\underline{\underline{A}}/A)}\right) \to \check{N} = \frac{1}{\eta\left[1 - \mathrm{E}(\underline{\underline{A}}/A)\right]}. \tag{A.3}$$

In equilibrium, under low barriers to entry, we have $N > \check{N}$ and, consequently, some low-productivity firms remain inactive—that is, $\underline{A} > \underline{\underline{A}}$. This implies that $s(\underline{A}) = 0$, as formally stated in the next proposition:

Proposition A.2 *For $N > \check{N}$, $s(\underline{A}) = 0$.*

Proof. It is sufficient to show that, for $N > \check{N}$, there exists a productivity cutoff $\underline{A} \in [\underline{\underline{A}}, \overline{A}]$ such that $s(\underline{A}) = 0$; after all, in that case, neither $s(\underline{A}) > 0$ nor $s(\underline{A}) < 0$ could be an equilibrium. The existence of such $\underline{A}$ follows from three results: (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 (a) $N > \check{N}$ implies not all firms can be active simultaneously, and (b) $s(A)$ is increasing in $A \in [\underline{A}, \overline{A}]$. The second result follows directly from Equation (A.2): as $\underline{A} \to \overline{A}$, we have $s(\underline{A}) \to 1/N_a = 1/[N(1 - G(\underline{A}))] \to +\infty$. The third result holds because both $N_a \equiv N(1 - G(\underline{A}))$ and $\mathrm{E}_a(1/A)$ are continuous in $\underline{A}$; the continuity of the latter stems from its differentiability in $\underline{A}$, as established by Proposition A.1. ■

Finally, plugging $s(\underline{A}) = 0$ into Equations (A.1) and (A.2), we obtain the expressions shown in Section 2:

$$s(A_i) = \eta\left(1 - \underline{A}/A_i\right) \tag{7}$$

$$\eta = \frac{1}{N_a\left[1 - \mathrm{E}_a\left(\underline{A}/A\right)\right]}. \tag{8}$$

A.2 Entry and the productivity cutoff for active intermediate firms

Equation (8) implicitly defines the productivity cutoff $\underline{A}$ as a function of the total number of firms N. The next proposition shows that this function is well defined for $N \geq \check{N}$, with $\underline{A}$ increasing in N—that is, greater entry raises the cutoff $\underline{A}$.

Proposition A.3 *Equation* (8) *implicitly defines the productivity cutoff $\underline{A}$ as an increasing and well-defined function of the total number of firms N, for $N \geq \check{N}$. In particular, as $N \to +\infty$, we have $\underline{A} \to \overline{A}$.*

Proof. Equation (8) can be rewritten as $f(\underline{A}) = (\eta N)^{-1}$, where $f(\underline{A}) \equiv (1-G(\underline{A}))\left[1 - \mathrm{E}_a\left(\underline{A}/A\right)\right]$. The derivative of this function f is given by

$$\begin{aligned}\frac{\partial f(\underline{A})}{\partial \underline{A}} =& - g(\underline{A})\left[1 - \mathrm{E}_a\left(\underline{A}/A\right)\right] - (1 - G(\underline{A}))\left[\mathrm{E}_a\left(1/A\right) + \underline{A} g_a(\underline{A})\left(\mathrm{E}_a\left(1/A\right) - 1/\underline{A}\right)\right] \\ \frac{\partial f(\underline{A})}{\partial \underline{A}} =& - (1 - G(\underline{A}))\mathrm{E}_a\left(1/A\right) < 0,\end{aligned}$$

where the first line uses Proposition A.1 with $h(A) = 1/A$. Since f is decreasing in $\underline{A}$, any increase in N must be offset by a higher $\underline{A}$ to maintain Equation (8) in balance.

To show that a solution exists for all $N \geq \check{N}$, note first that the equality holds by construction for $N = \check{N}$ (see Equation (A.3)), when $\underline{A} = \underline{\underline{A}}$. For any $N > \check{N}$—and hence $\underline{A} > \underline{\underline{A}}$—a solution exists because (i) f is continuous in $\underline{A}$, as implied by its differentiability, and (ii) $f(\underline{A}) \to 0$ as $\underline{A} \to \overline{A}$, since both $G(\underline{A})$ and $\mathrm{E}_a\left(\underline{A}/A\right)$ converge to one in that limit.[29]

Finally, it remains to show that $\underline{A} \to \overline{A}$ as $N \to +\infty$. Note that in this case, $f(\underline{A}) \to 0$ to satisfy Equation (8). But we have just shown that this occurs precisely when $\underline{A} \to \overline{A}$. Then, the result follows directly, since $\underline{A}$ is a well-defined function of N. ■

While $\underline{A}$ increases with N, the relationship between N and N_a is less clear. From Proposition A.1, we have

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

which may be positive or negative depending on the distribution. What we can ensure is that $N_a \to +\infty$ as $N \to +\infty$; otherwise, Equation (8) would not hold, since $\mathrm{E}_a\left(\underline{A}/A\right) \to 1$ in this case.

A.3 SOC for profit maximization of intermediate firms

The profit of an intermediate firm i in sector s is given by $\pi_{si} = (p_s - MC_{si})\, Y_{si}$, where p_s is determined by the inverse demand function (2)—i.e., $p_s = (Y/Y_s)^{1/\eta}$. Consequently,

$$\frac{\partial \pi_{si}}{\partial Y_{si}} = - (1/\eta) p_s s_{si} + (p_s - MC_{si}) \tag{A.4}$$

$$\frac{\partial^2 \pi_{si}}{\partial^2 Y_{si}} = (1/\eta)(1/\eta + 1) p_s Y_s^{-1} s_{si} - 2(1/\eta) p_s Y_s^{-1}, \tag{A.5}$$

which implies the following second-order condition (SOC):

$$s_{si} < \frac{2\eta}{1+\eta}. \tag{A.6}$$

[29] To see that $\mathrm{E}_a\left(\underline{A}/A\right) \to 1$ as $\underline{A} \to \overline{A}$, note $\underline{A}/\overline{A} = \mathrm{E}_a\left(\underline{A}/\overline{A}\right) \leq \mathrm{E}_a\left(\underline{A}/A\right) \leq \mathrm{E}_a\left(\underline{A}/\underline{A}\right) = 1$.

To assess this condition, we consider the model solution for the market share of an active firm (Equation (7)), derived in Appendix A.1. Using this expression, we show that the SOC holds for low $\eta > 0$, consistent with the assumption used to ensure a high number of active firms. The next proposition states this low-η condition more precisely, where, once again, we drop sectoral subscripts since sectors are equivalent in equilibrium.

Proposition A.4 *The SOC for profit maximization of intermediate firms holds if* $\eta \in \left(0, \frac{1+\underline{\underline{A}}/\overline{A}}{1-\underline{\underline{A}}/\overline{A}}\right]$.

Proof. To ensure the SOC (A.6) is satisfied, we require $\frac{2\eta}{1+\eta} > s_i$. From Equation (7), we have $s_i < \eta\left(1 - \underline{\underline{A}}/\overline{A}\right)$, which is a strict inequality since $\underline{A} > \underline{\underline{A}}$ due to $N > \check{N}$. As a result, the SOC holds for every active firm if

$$\frac{2\eta}{1+\eta} \geq \eta\left(1 - \underline{\underline{A}}/\overline{A}\right) \rightarrow \eta \leq \frac{1+\underline{\underline{A}}/\overline{A}}{1-\underline{\underline{A}}/\overline{A}}.$$

Since $\eta > 0$ by assumption, the result follows. ■

It is worth noting that $\frac{1+\underline{\underline{A}}/\overline{A}}{1-\underline{\underline{A}}/\overline{A}} > 1$; therefore, the SOC always holds for $\eta \in (0, 1]$. For $\eta > 1$, the SOC is satisfied only if the elasticity is not too large—that is, only if $\eta \in \left(1, \frac{1+\underline{\underline{A}}/\overline{A}}{1-\underline{\underline{A}}/\overline{A}}\right)$. This result may seem surprising, as in oligopoly models the condition $\eta > 1$ is typically sufficient for the SOC to hold. The key difference here lies in the assumption of a continuum of firms. In the discrete case, market shares always satisfy $s_i \in (0, 1)$, which ensures that the SOC holds for $\eta > 1$, since $\frac{2\eta}{1+\eta} > 1$ in this case. However, in the continuous case, s_i behaves like a density function and may exceed one. In fact, one can show that, for a given N, $s(\overline{A}) \rightarrow +\infty$ as $\eta \rightarrow +\infty$. To see this, note from Equation (8) that $\underline{A} \rightarrow \overline{A}$ as $\eta \rightarrow +\infty$, which implies that $\eta \mathrm{E}_a(1 - \underline{A}/A) = \frac{1}{N_a} = \frac{1}{N(1-G(\underline{A}))} \rightarrow +\infty$. Since $(1 - \underline{A}/\overline{A}) \geq \mathrm{E}_a(1 - \underline{A}/A)$, it follows that $s(\overline{A}) = \eta(1 - \underline{A}/\overline{A}) \rightarrow +\infty$ as $\eta \rightarrow +\infty$.

A.4 Optimal use of inputs by intermediate firms

Intermediate firms use inputs optimally, taking the same rental prices as given. As a consequence, from the FOCs of the cost minimization problem of an active firm i, we have

$$\frac{w}{MPL_i} = \frac{r}{MPK_i} \leftrightarrow \frac{wL_i}{(1-\alpha)Y_i} = \frac{rK_i}{\alpha Y_i} \leftrightarrow \frac{K_i}{L_i} = \frac{w}{r}\frac{\alpha}{1-\alpha}, \tag{A.7}$$

where MPL_i and MPK_i denote the marginal products of labor and capital for firm i, respectively. Thus, every active firm chooses the same capital–labor ratio. Consequently, from the above expression, for any firm j and active firm i, $K_j = \frac{K_i}{L_i} L_j$, which implies

$$\theta_{Ki} \equiv \frac{K_i}{K} = \frac{K_i}{\int_0^N K_j dj} = \frac{K_i}{\int_0^N \frac{K_i}{L_i} L_j dj} = \frac{L_i}{\int_0^N L_j dj} = \frac{L_i}{L} \equiv \theta_{Li}. \tag{A.8}$$

This allows us to define $\theta_i \equiv \theta_{Li} = \theta_{Ki}$ for every firm i. Given this, we can rewrite Equation (A.7) as

$$\alpha = \frac{K_i r}{K_i r + L_i w} = \frac{\theta_i K r}{\theta_i K r + \theta_i L w} = \frac{Kr}{Kr + Lw}, \tag{A.9}$$

meaning α equals the cost share of capital.

A.5 Allocative efficiency: derivation and properties

From Equations (4) and (5), we obtain:

$$Y = \int_0^N A_i K_i^{\alpha} L_i^{1-\alpha} di = \overline{A}\Omega K^{\alpha} L^{1-\alpha}, \tag{9}$$

where $K \equiv \int_0^N K_i di$, $L \equiv \int_0^N L_i di$, and $\Omega \equiv \int_0^N \theta_{Ki}^{\alpha}\theta_{Li}^{1-\alpha}\left(A_i/\overline{A}\right) di$.

Moreover,

$$s(A_i) = \frac{Y_i}{Y} = \frac{A_i K_i^{\alpha} L_i^{1-\alpha}}{\overline{A}\Omega K^{\alpha} L^{1-\alpha}} = \frac{A_i \theta_i}{\overline{A}\Omega} \rightarrow \theta_i = \overline{A}\Omega \frac{s(A_i)}{A_i} \tag{A.10}$$

$$\overline{A}\Omega = \frac{1}{\int_0^N \frac{s(A_i)}{A_i} di}, \tag{A.11}$$

where we use (5), (9), and $\theta_{Ki} = \theta_{Li} \equiv \theta_i$ (Equation (A.8)) to get (A.10), while (A.11) is obtained by summing (A.10) over all firms. Therefore, as in Edmond et al. (2015, 2023), aggregate productivity is a quantity-weighted harmonic mean of firms' productivity.

Finally, substituting Equations (7) and (8) into (A.11), we obtain the expression presented in Section 2:

$$\Omega = \frac{\mathrm{E}_a\left[(\underline{A}/\overline{A})(1 - \underline{A}/A)\right]}{\mathrm{E}_a\left[(\underline{A}/A)(1 - \underline{A}/A)\right]}. \tag{10}$$

Using these results, we demonstrate three important properties of allocative efficiency, as summarized in the following proposition:

Proposition A.5 *Allocative efficiency Ω satisfies (i) $\Omega \in (0, 1]$; (ii) $\Omega \rightarrow 1$ as $\underline{A} \rightarrow \overline{A}$; and (iii) Ω is increasing in $\underline{A}$ as $\frac{\partial \Omega}{\partial \underline{A}} > 0$.*

Proof. (i) That $\Omega \in (0, 1]$ follows directly from Equation (10), given that $A \in [\underline{A}, \overline{A}]$ for all active firms. Alternatively, we can use the definition $\Omega \equiv \int_0^N \theta_{Ki}^{\alpha}\theta_{Li}^{1-\alpha}\left(A_i/\overline{A}\right) di$, since it implies $\Omega > 0$ and $\Omega \leq \int_0^N \theta_{Ki}^{\alpha}\theta_{Li}^{1-\alpha} di \leq 1$. To verify this last inequality, use the fact that $\theta_{Ki}^{\alpha}\theta_{Li}^{1-\alpha} = \exp\left(\alpha \ln(\theta_{Ki}) + (1-\alpha)\ln(\theta_{Li})\right) \leq \exp\left[\ln(\alpha\theta_{Ki} + (1-\alpha)\theta_{Li})\right] = \alpha\theta_{Ki} + (1-\alpha)\theta_{Li}$, applying Jensen's inequality (since $\ln$ is concave) and noting that $\exp$ is increasing.

(ii) To establish that $\Omega \to 1$ as $\underline{A} \to \overline{A}$, note from Equation (A.11) that

$$1 = \int_0^N s(A_i)(\overline{A}/\overline{A})di \leq \Omega^{-1} \leq \int_0^N s(A_i)(\overline{A}/\underline{A})di = \overline{A}/\underline{A},$$

which implies $1 \leq \lim_{\underline{A}\to\overline{A}} \Omega^{-1} \leq \lim_{\underline{A}\to\overline{A}} \overline{A}/\underline{A} = 1$. Therefore, $\lim_{\underline{A}\to\overline{A}} \Omega = 1$.

(iii) To show that Ω is increasing in $\underline{A}$, differentiate Equation (10) with respect to $\underline{A}$:

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

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

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

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

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

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

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 A.1 to obtain expressions for $\frac{\partial \mathrm{E}_a(1/A)}{\partial \underline{A}}$ and $\frac{\partial \mathrm{E}_a(1/A^2)}{\partial \underline{A}}$ in the fourth line. Thus, $\frac{\partial \Omega}{\partial \underline{A}} > 0$. ■

It is worth mentioning that the first two results are not specific to the Cournot model; they rely essentially on the assumption of homogeneous goods. First, the property $\Omega \in (0, 1]$ follows directly from the fact that Ω measures the distance between actual TFP, $\overline{A}\Omega$, and its optimal level, $\overline{A}$. This fact depends solely on the assumption of homogeneous goods, which is precisely what allows us to derive this property from the definition of Ω. Second, we showed that $\Omega \to 1$ as $\underline{A} \to \overline{A}$ using Equation (A.11), which requires only (i) homogeneous goods and (ii) optimal input use. Since the optimal allocation is independent of the specific form of firm behavior, this result does not even require optimal input use; it is sufficient that firms with the same productivity level behave identically. In this case, for any active firm i, we can write $\theta_{Ki} = \theta_K(A_i)$ and $\theta_{Li} = \theta_L(A_i)$, which implies $\theta_{Ki} = 1/N_a = \theta_{Li}$ if all firms are equally productive. Under these conditions, the second property of Ω follows directly from its definition, $\Omega \equiv \int_0^N \theta_{Ki}^{\alpha}\theta_{Li}^{1-\alpha}\left(A_i/\overline{A}\right) di = N_a \int_{\underline{A}}^{\overline{A}} \theta_K(A)^{\alpha}\theta_L(A)^{1-\alpha}\left(A/\overline{A}\right) g_a(A)dA$, since $\Omega \to N_a \int_{\underline{A}}^{\overline{A}}(1/N_a)\left(\overline{A}/\overline{A}\right) g_a(A)dA = 1$ as $\underline{A} \to \overline{A}$.

A.6 Derivation of the cost-weighted average markup

Recall from Section 2 that

$$\mu(A_i) \equiv \mu_i = A_i/\underline{A}, \tag{11}$$

while the cost-weighted average markup is defined as $\mu \equiv \int_0^N \left(\frac{L_i w + K_i r}{Lw + Kr}\right) \mu_i di = \int_0^N \theta_i \mu_i di$. Substituting Equations (A.10) and (11) into this expression yields

$$\mu = \int_0^N \left[\overline{A}\Omega \frac{s(A_i)}{A_i}\right] \left(\frac{A_i}{\underline{A}}\right) di = \frac{\overline{A}\Omega}{\underline{A}}. \tag{12}$$

As Edmond et al. (2023) point out, when firms share the same returns to scale, the cost-weighted *arithmetic* average markup is equivalent to the sales-weighted *harmonic* average markup. Since, in our model, all firms operate under constant returns to scale, this equivalence also holds here. To see this, use Equations (A.10) and (11) to obtain

$$\theta_i = \frac{\overline{A}\Omega}{\underline{A}} \frac{s_i}{\mu_i} \quad \Longrightarrow \quad \frac{s_i}{\mu_i} = \left(\frac{\overline{A}\Omega}{\underline{A}}\right)^{-1} \theta_i,$$

which then implies

$$\left(\int_0^N \frac{s_i}{\mu_i} di\right)^{-1} = \frac{\overline{A}\Omega}{\underline{A}} \left(\int_0^N \theta_i di\right)^{-1} = \frac{\overline{A}\Omega}{\underline{A}} = \mu.$$

A.7 Expected profit of an entrant: derivation and properties

As argued in Section 2, the expected profit of an entrant in our model equals the average profit of incumbent firms, $\bar{\pi}$. Recalling that $p_s = 1$ and $Y_s = Y$ for all sectors s, we have:

$$\begin{aligned}
\bar{\pi} =& G(\underline{A}) \times 0 + (1 - G(\underline{A})) \times \mathrm{E}_a \left[(1 - 1/\mu(A))\, s(A) Y \right] \\
\bar{\pi} =& (1 - G(\underline{A})) \left(\overline{A} \Omega K^{\alpha} L^{1-\alpha} \right) \eta \mathrm{E}_a \left[(1 - \underline{A}/A)^2 \right] \\
\bar{\pi} =& K^{\alpha} L^{1-\alpha} \left[\eta (1 - G(\underline{A})) \mathrm{E}_a \left(1 - \underline{A}/A\right) \right] \overline{A} \Omega \mathrm{E}_a \left[\frac{(1 - \underline{A}/A) - (\underline{A}/A)\left(1 - \underline{A}/A\right)}{\mathrm{E}_a \left(1 - \underline{A}/A\right)} \right] \\
\bar{\pi} =& K^{\alpha} L^{1-\alpha} \left(\frac{\overline{A}\Omega - \underline{A}}{N} \right), \qquad (13)
\end{aligned}$$

where we use, in the second line: (i) $Y = \overline{A} \Omega K^{\alpha} L^{1-\alpha}$ (Equation (9)), (ii) $\mu(A) = A/\underline{A}$ (Equation (11)), and (iii) $s(A) = \eta \left(1 - \underline{A}/A\right)$ (Equation (7)); and in the final line: (i) $N^{-1} = \eta(1 - G(\underline{A})) \left[1 - \mathrm{E}_a \left(\underline{A}/A\right)\right]$ (from Equation (8)), and (ii) $\Omega = \frac{\mathrm{E}_a \left[(\underline{A}/\overline{A})(1 - \underline{A}/A) \right]}{\mathrm{E}_a \left[(\underline{A}/A)(1 - \underline{A}/A) \right]}$ (Equation (10)).

Equation (13) implies that $\bar{\pi}$ may be non-monotonic in N because both $\underline{A}$ and Ω increase with N (Propositions A.3 and A.5).[30] In words, while a higher N exerts downward pressure on expected profits through standard competition effects, it also improves allocative efficiency, which raises sectoral output—i.e., market size—and thereby positively affects expected profits. Figure A.1 illustrates this possibility using a uniform productivity distribution, where we normalize $\underline{\underline{A}} = K^{\alpha} L^{1-\alpha} = \eta = 1$ and set $\overline{A} = 30$. This non-monotonicity appears only under extreme scenarios characterized by very high productivity dispersion among active firms. In this simulation, $\bar{\pi}$ is increasing only for N close to $\check{N}$, when the productivity cutoff $\underline{A}$ is near $\underline{\underline{A}}$, the cost-weighted average markup μ exceeds 8, and allocative efficiency Ω is below 0.4.[31] Intuitively, when misallocation is severe, a marginal increase in N can yield a sufficiently large improvement in allocative efficiency to offset the effect of intensified competition, thereby raising $\bar{\pi}$.

Hence, the monotonicity of $\bar{\pi}$ cannot be guaranteed. However, we can establish two key properties of this function, as stated in the following proposition:

[30] The potential non-monotonicity of $\bar{\pi}$ stems partly from our assumption that entry costs are purely pecuniary and do not absorb real resources (e.g., labor or capital). As a result, increased entry expands aggregate output by reducing misallocation without crowding out production inputs—thereby isolating the pure effect of misallocation on entry and output. Introducing an entry cost denominated in labor, as in Edmond et al. (2023), would strengthen the negative impact of entry on expected profits. However, such a specification does not guarantee the monotonicity of $\bar{\pi}$; the outcome would ultimately depend on the parameter calibration.

[31] Empirical evidence from firm-level data indicates that cost-weighted average markups μ above 8 are extremely rare. See De Loecker et al. (2020) for the US (1955–2016), Traina (2018) for the US (1950–2016), Baqaee and Farhi (2020) for the US (1985–2015), and De Loecker and Eeckhout (2018) for 134 countries between 1980 and 2016. For instance, De Loecker and Eeckhout (2018) find that global cost-weighted average markups are always below 1.6. The sales-weighted average markup—which tends to overstate μ (Edmond et al., 2023)—remains below 2 across global regions (Europe, North America, South America, Asia, Oceania, and Africa) and below 4 for the 40 countries presented in their Appendix A.

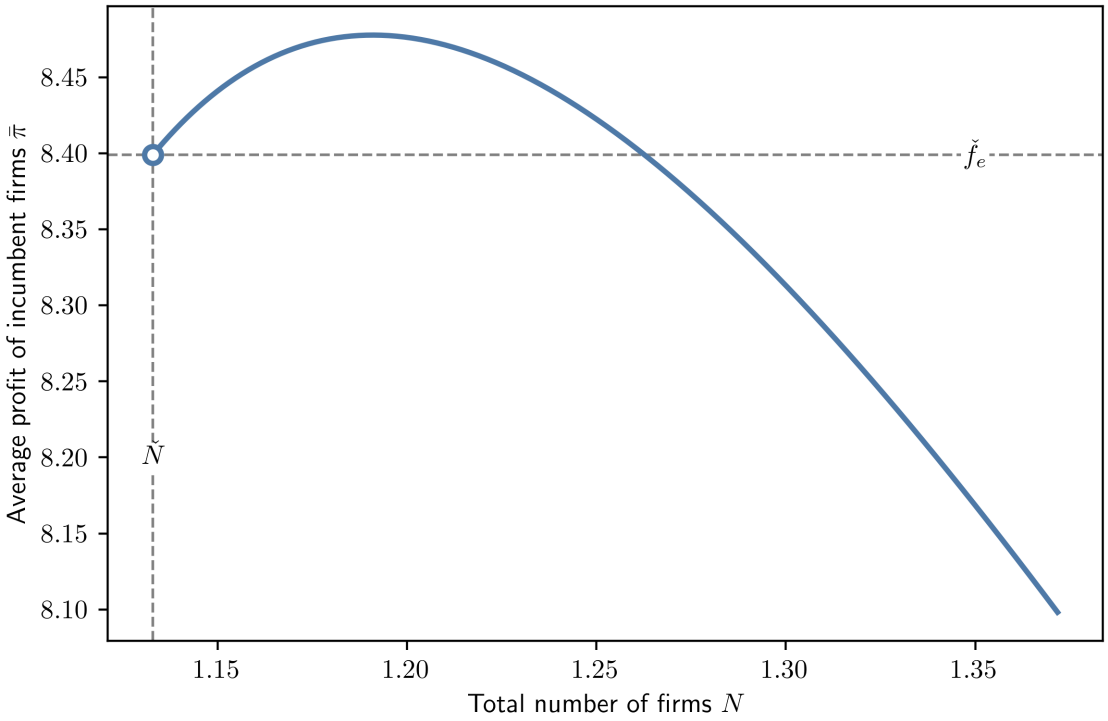


Figure A.1: Average profit of incumbent firms $\bar{\pi}$ *versus* total number of firms N.

Proposition A.6 *The average profit of incumbent firms, $\bar{\pi}$, satisfies: (i) $\bar{\pi}$ is continuous in N, and (ii) as $N \to +\infty$—i.e., in the limit of perfect competition—$\bar{\pi} \to 0$.*

Proof. (i) The continuity of $\bar{\pi}$ follows from the fact that, in Equation (13), it is a composition of continuous functions of N: namely, $\underline{A}$, Ω, and N itself. To see that $\underline{A}$ and Ω are continuous in N, note that Ω is differentiable in $\underline{A}$ and $\underline{A}$ is differentiable in N, as shown in the proofs of Propositions A.5 and A.3, respectively.

(ii) As $N \to +\infty$, Proposition A.3 implies that $\underline{A} \to \overline{A}$, which in turn leads to $\Omega \to 1$, as established in Proposition A.5. Consequently, $\overline{A}\Omega - \underline{A} \to \overline{A} - \overline{A} = 0$ as $N \to +\infty$, which yields the stated result via Equation (13). ■

A.8 Free-entry equilibrium: the general case

If $\bar{\pi}$ decreases with N—as assumed in the main text—it follows directly that the equilibrium condition $f_e = \bar{\pi}$ implicitly defines N as a decreasing function of f_e. However, as discussed in Appendix A.7, the monotonicity of $\bar{\pi}$ cannot be guaranteed, so the values of N satisfying $f_e = \bar{\pi}$ may not be unique. We address this possible multiplicity by selecting the largest such N—a choice consistent with the notion of a free-entry equilibrium, as it ensures that no profitable entry opportunities remain. Interestingly, this procedure not only defines N as a well-defined function of f_e but also ensures that this function is decreasing, despite the potentially ambiguous

effect of N on $\bar{\pi}$. Thus, the result is not specific to the monotonic case; it holds in general, even when $\bar{\pi}$ is non-monotonic.

This result is formalized in the next proposition. Before proceeding, recall from Section 2 that $\check{f}_e$ denotes the average profit of incumbent firms when $N = \check{N}$ and, therefore, $\underline{A} = \underline{\overline{A}}$. Formally, from Equation (13):

$$\check{f}_e = K^{\alpha} L^{1-\alpha} \left(\frac{\overline{A}\check{\Omega} - \underline{\overline{A}}}{\check{N}} \right), \tag{A.12}$$

where $\check{\Omega} \equiv \frac{\mathrm{E}\left[(\underline{A}/\overline{A})(1-\underline{A}/A)\right]}{\mathrm{E}\left[(\underline{A}/A)(1-\underline{A}/A)\right]}$.

Proposition A.7 *By selecting the largest total number of firms N that satisfies the condition $f_e = \bar{\pi}$, we implicitly define N as a well-defined and decreasing function of f_e, where this function $N : (0, \check{f}_e) \to (\check{N}, +\infty)$. In particular, $N \to +\infty$ as $f_e \to 0$.*

Proof. By construction, $\check{N}$ is a solution to $\check{f}_e = \bar{\pi}$. Therefore, for any $f_e \in (0, \check{f}_e)$, the equation $f_e = \bar{\pi}$ admits at least one solution, since (i) $\bar{\pi}$ is continuous in N, and (ii) $\bar{\pi} \to 0$ as $N \to +\infty$ (by Proposition A.6). If the solution is not unique, we select the largest such N, ensuring that N is a well-defined function of f_e. Furthermore, this function is decreasing: a lower f_e must correspond to a higher N; otherwise, it would imply that the lower f_e fails to intersect $\bar{\pi}$ for any larger N, contradicting the properties of $\bar{\pi}$ stated in Proposition A.6.

Finally, the codomain of this function is $(\check{N}, +\infty)$—i.e., $N(f_e) > \check{N}$ for any $f_e \in (0, \check{f}_e)$—because (i) $N(\check{f}_e) \geq \check{N}$—holding with strict equality if the solution to $\check{f}_e = \bar{\pi}$ is unique—and (ii) $N(f_e)$ decreases with f_e, as just established. Moreover, since $\bar{\pi} \to 0$ when $N \to +\infty$ (Proposition A.6), it follows directly that $N \to +\infty$ as $f_e \to 0$, given that we select the largest N satisfying $f_e = \bar{\pi}$. ■

A final comment is in order. In the main text, we assumed for simplicity that $\bar{\pi}$ decreases with N, in which case $N > \check{N}$—as implied by our low-barriers assumption—if and only if $f_e < \check{f}_e$. In the general, possibly non-monotonic, case, the condition $f_e < \check{f}_e$ remains sufficient to guarantee that $N > \check{N}$, since (i) $N(\check{f}_e) \geq \check{N}$, as $N(\check{f}_e)$ is the largest N satisfying $\check{f}_e = \bar{\pi}$—which necessarily holds for $N = \check{N}$ by construction, and thus $N(\check{f}_e) = \check{N}$ in the monotonic case, when the solution to $\check{f}_e = \bar{\pi}$ is unique—and (ii) $N(f_e)$ decreases with f_e, as established in Proposition A.7. However, this condition is not necessary, since an f_e greater than $\check{f}_e$ may still imply $N > \check{N}$. This result is illustrated by the exercise with a uniform distribution shown in Figure A.1.

B Calibrating the productivity bounds among active firms: conditions for exact moment matching

Given the shape of the productivity distribution, determining the productivity bounds among active firms—$\underline{A}$ and $\overline{A}$—is sufficient to compute the key model outcomes, such as allocative efficiency Ω. Section 3.1 proposes calibrating these bounds by targeting (i) aggregate TFP, $\overline{A}\Omega$, and (ii) the cost-weighted average markup, μ. These two target moments directly pin down the lower bound, since Equation (12) implies $\underline{A} = \overline{A}\Omega/\mu$.

The difficulty, then, is to recover the upper bound $\overline{A}$, since it requires solving

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

This appendix studies the existence and uniqueness of a solution to this equation. Formally, we show that under certain conditions, Equation (14) implicitly defines $\overline{A}$ as a well-defined function of μ. This result—made possible by the tractability of our model—guarantees an exact match of both target moments.

We proceed by first deriving conditions that apply to any continuous truncated distribution defined for any $\overline{A} \in (\underline{A}, +\infty)$. We then specialize this general framework to the case of a Pareto distribution with shape parameter $k \neq 0$. Within the Pareto case, we also examine the limiting behavior of allocative efficiency Ω as $k \to -\infty$.

B.1 Arbitrary distribution

Assume that $A \in [\underline{\underline{A}}, \overline{A}]$ is distributed according to a density function g, with cumulative distribution function G, for any $\overline{A}$ satisfying $0 < \underline{\underline{A}} < \overline{A} < +\infty$. Then, for $\underline{A} \in (\underline{\underline{A}}, \overline{A})$, the density over the interval $[\underline{A}, \overline{A}]$ is given by $g_a(A) \equiv \frac{g(A)}{1-G(\underline{A})} > 0$, with associated cumulative distribution function G_a.

Since both g and g_a are defined for any $\overline{A} \in (\underline{A}, +\infty)$, we further introduce a density $\hat{g}$ and its corresponding cumulative distribution function $\hat{G}$ over the extended support $A \in \left[\underline{A}, \overline{\overline{A}}\right]$, for some $\overline{\overline{A}} > \overline{A}$, such that $g_a(A) = \frac{\hat{g}(A)}{\hat{G}(\overline{A})}$.

Proposition B.1 *Let h be a function of A such that (i) $\frac{\partial h(A)}{\partial A} = 0$ and (ii) $\mathrm{E}_a\left(h(A)\right)$ is well defined. Then,*

$$\frac{\partial \mathrm{E}_a\left(h(A)\right)}{\partial \overline{A}} = g_a(\overline{A})\left[h(\overline{A}) - \mathrm{E}_a\left(h(A)\right)\right].$$

Proof. Given a function h satisfying the two conditions in the statement,

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

where we use the fact that $\frac{\partial\left[\int_{\underline{A}}^{\overline{A}} h(A)\hat{g}(A)dA\right]}{\partial \overline{A}} = h(\overline{A})\hat{g}(\overline{A})$, since h does not depend on $\overline{A}$. ■

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

Proof. Differentiating Equation (14) with respect to $\overline{A}$ yields:

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

where the third line uses Proposition B.1 to compute $\frac{\partial \mathrm{E}_a(\underline{A}/A)}{\partial \overline{A}}$ and $\frac{\partial \mathrm{E}_a\left[(\underline{A}/A)^2\right]}{\partial \overline{A}}$. The final expression is positive because $(1-\underline{A}/A)(\underline{A}/A - \underline{A}/\overline{A}) > 0$ for all $A \in (\underline{A}, \overline{A})$. This completes the proof. ■

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

Proof. From Equation (14), we can bound μ as follows:

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

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

Proposition B.4 *Equation* (14) *implicitly defines the technology frontier* $\overline{A} \in (\underline{A}, +\infty)$ *as a continuous, increasing, and well-defined function of the cost-weighted average markup* μ*, for* $\mu \in \left(1, \lim_{\overline{A}\to+\infty}\mu\right)$*. However, if* $\mu \geq \lim_{\overline{A}\to+\infty}\mu$*, there exists no* $\overline{A} \in (\underline{A}, +\infty)$ *that solves Equation* (14)*.*

Proof. First, μ is continuous and increasing in $\overline{A} \in (\underline{A}, +\infty)$, as it is differentiable with $\frac{\partial \mu}{\partial \overline{A}} > 0$ (Proposition B.2). This implies that μ is a one-to-one function of $\overline{A}$. Second, by Proposition B.3, we have $\lim_{\overline{A} \to \underline{A}^+} \mu = 1$. Thus, since μ is continuous and increasing in $\overline{A}$, its image is $\left(1, \lim_{\overline{A} \to +\infty} \mu\right)$. As a consequence of these two results, the inverse mapping from μ to $\overline{A}$ is also continuous, increasing, and well-defined for $\mu \in \left(1, \lim_{\overline{A} \to +\infty} \mu\right)$, with $\overline{A} \in (\underline{A}, +\infty)$. Moreover, the same two results imply that no $\overline{A} \in (\underline{A}, +\infty)$ solves Equation (14) for any $\mu \geq \lim_{\overline{A} \to +\infty} \mu$. ■

B.2 Pareto distribution of firm productivity

Assume that $A \in \left[\underline{\underline{A}}, \overline{A}\right] \subset (0, +\infty)$ follows a truncated Pareto distribution with shape parameter k, implying

$$g(A) = k \left(\frac{\underline{\underline{A}}^k \overline{A}^k}{\overline{A}^k - \underline{\underline{A}}^k} \right) A^{-k-1} \tag{B.1}$$

$$G(A) = \frac{\overline{A}^k - \underline{\underline{A}}^k \overline{A}^k A^{-k}}{\overline{A}^k - \underline{\underline{A}}^k}. \tag{B.2}$$

This distribution is well defined for any $k \neq 0$, as $g(A) > 0$ for all $A \in \left[\underline{\underline{A}}, \overline{A}\right]$ and $G(\overline{A}) = 1$. In particular, when $k = -1$, it corresponds to the uniform distribution.

It is straightforward to verify that $A \in \left[\underline{A}, \overline{A}\right]$ also follows a truncated Pareto distribution with the same shape parameter k, since its density is

$$g_a(A) \equiv \frac{g(A)}{1 - G(\underline{A})} = k \left(\frac{\underline{A}^k \overline{A}^k}{\overline{A}^k - \underline{A}^k} \right) A^{-k-1} = k \left(\frac{\overline{A}^k}{\tilde{A}^k - 1} \right) A^{-k-1}, \tag{B.3}$$

where $\tilde{A} \equiv \overline{A}/\underline{A} > 1$.

Proposition B.5 *Let $A \in \left[\underline{\underline{A}}, \overline{A}\right]$ follow a truncated Pareto distribution with shape parameter k, with $k \neq 0$. Then, for any integer $j \geq 1$,*

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

Proof. Let $k \neq 0$ and $j \in \mathbb{N} \setminus \{0\}$. From the truncated Pareto density (B.3), if $k + j \neq 0$,

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

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

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

Proposition B.6 *Let $A\in\left[\underline{A},\overline{A}\right]$ follow a truncated Pareto distribution with shape parameter k, with $k\neq 0$. Then,*

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

Proof. If $k=-1$, using Proposition B.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 B.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 $k\neq 0,-1,-2$,

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

where we once again use Proposition B.5 and Equation (14). ■

Proposition B.7 *Let $A\in\left[\underline{A},\overline{A}\right]$ follow a truncated Pareto distribution with shape parameter k, with $k\neq 0$. Then,*

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

Proof. Initially, note that $\lim_{\overline{A}\to+\infty}\mu = \lim_{\tilde{A}\to+\infty}\mu$, since μ is solely a function of $\tilde{A} \equiv \overline{A}/\underline{A}$ (see Proposition B.6). Thus, it is sufficient to compute $\lim_{\tilde{A}\to+\infty}\mu$.

If $k > 0$—so that $k + j \neq 0$ for any integer $j \geq 1$—then Proposition B.5 implies

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

which, combined with Equation (14), yields

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

If $k < 0$, Proposition B.5 gives the following limits, 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. As a result, from Equation (14), $\lim_{\tilde{A}\to+\infty}\mu = +\infty$ when $k < 0$. ■

Proposition B.8 *Let $A \in \left[\underline{A}, \overline{A}\right]$ follow a truncated Pareto distribution with shape parameter k, with $k \neq 0$. Then, Equation* (14) *implicitly defines the technology dispersion $\tilde{A} \equiv (\overline{A}/\underline{A}) \in (1, +\infty)$ as a continuous, increasing, and well-defined function of the cost-weighted average markup μ, for all μ satisfying (i) $\mu > 1$, and (ii) $k < \frac{2}{\mu-1}$. However, if $\mu > 1$ but $k \geq \frac{2}{\mu-1}$—a case that can arise only when $k > 0$—then no $\tilde{A} \in (1, +\infty)$ solves Equation* (14)*.*

Proof. From Proposition B.4, Equation (14) implicitly defines $\overline{A} \in (\underline{A}, +\infty)$ as a continuous, increasing, and well-defined function of μ, for $\mu \in \left(1, \lim_{\overline{A}\to+\infty}\mu\right)$. However, if $\mu \geq \lim_{\overline{A}\to+\infty}\mu$, no $\overline{A} \in (\underline{A}, +\infty)$ solves Equation (14). From Proposition B.6, in the Pareto case, μ depends only on $\tilde{A} \equiv \overline{A}/\underline{A}$ (given $k \neq 0$), so the same properties carry over to $\tilde{A} \in (1, +\infty)$.

Moreover, using Proposition B.7, the condition $\mu \in \left(1, \lim_{\overline{A}\to+\infty}\mu\right)$ can be rewritten as: (i) $\mu > 1$ if $k < 0$, and (ii) $1 < \mu < \frac{k+2}{k}$ if $k > 0$. Note that $1 < \mu < \frac{k+2}{k}$ implies $k < \frac{2}{\mu-1}$, which is automatically satisfied when $k < 0$ and $\mu > 1$. Therefore, for any $k \neq 0$, the condition $\mu \in \left(1, \lim_{\overline{A}\to+\infty}\mu\right)$ holds if and only if $\mu > 1$ and $k < \frac{2}{\mu-1}$. Consequently, $\mu \geq \lim_{\overline{A}\to+\infty}\mu > 1$ arises if and only $k \geq \frac{2}{\mu-1} > 0$. ■

Figure B.1 illustrates the results of Proposition B.8, plotting $\tilde{A}$ against μ for truncated Pareto distributions with $k = 3$, 5, and 9, as well as for the uniform distribution (corresponding to $k = -1$). It is worth noting that the condition $k < \frac{2}{\mu-1}$ always holds if $k < 0$ and $\mu > 1$. Thus,

the function $\tilde{A}(\mu)$ is well defined for any $\mu > 1$ when $k < 0$—a result illustrated in Figure B.1 for the uniform case ($k = -1$).

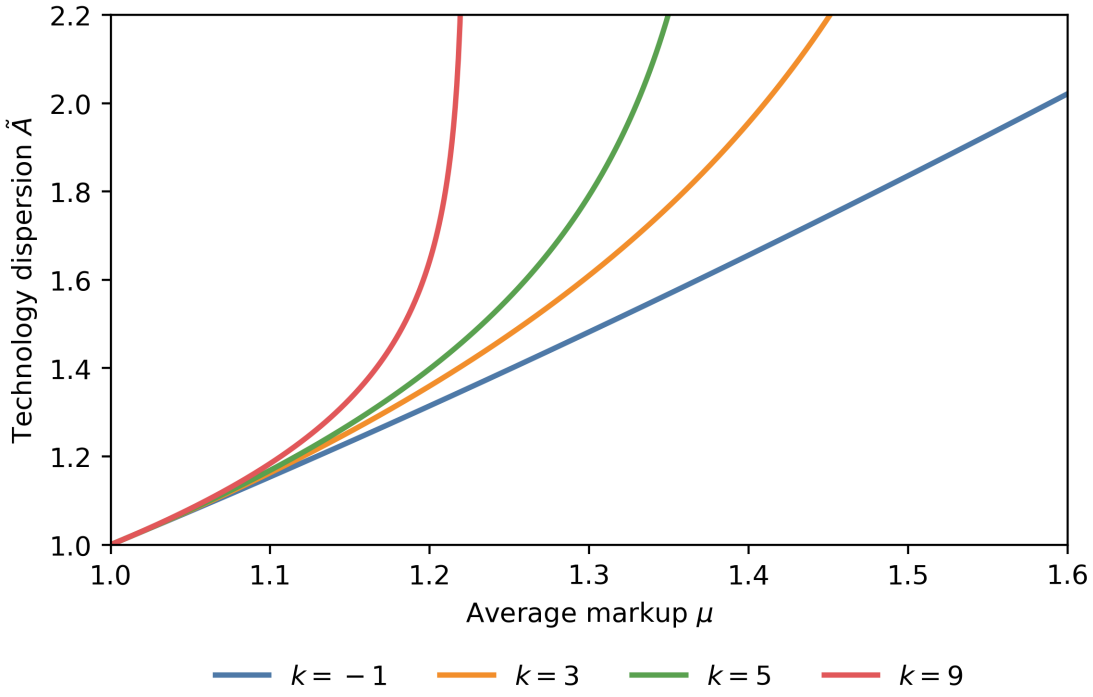


Figure B.1: Technological dispersion $\tilde{A} \equiv \overline{A}/\underline{A}$ *versus* average markup μ.

This ensures the feasibility of calibrating the productivity bounds among active firms even for highly negative values of k. However, an important question concerns the behavior of allocative efficiency Ω under these extreme conditions—a topic addressed in the concluding proposition of this section:

Proposition B.9 *Let $A \in [\underline{A}, \overline{A}]$ follow a truncated Pareto distribution with shape parameter k, with $k \neq 0$. Then,* $\lim_{k \to -\infty} \Omega = 1$.

Proof. From Proposition B.5, for any given integer $j \geq 1$, we have:

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

Consequently, it follows from Equation (10) that

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

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

C Model extension: beyond Cobb-Douglas production functions

In this appendix, we show that the model can be readily generalized by allowing intermediate firms to operate under an arbitrary well-behaved production function with $M > 0$ factors of production, provided that it exhibits (i) constant returns to scale and (ii) Hicks-neutral productivity. Formally, the production of firm i in sector s is now given by

$$Y_{si} = A_{si} f(F_{si}^1, \ldots, F_{si}^M), \tag{5'}$$

where $F_{si}^j \geq 0$ denotes the quantity of factor j used by firm i in sector s. The function f is assumed to be (i) homogeneous of degree one, (ii) differentiable, and (iii) well behaved, in the sense that the firm's cost-minimization problem admits a unique interior solution for positive rental prices.

We discuss the derivation of the model under the general production function (5') next. Before proceeding, however, it is useful to establish a preliminary result. Recall that intermediate firms choose inputs optimally, taking the same rental prices as given. As a consequence, the FOCs of the cost-minimization problem for a firm i in a sector s imply that, for any pair of factors r and k,

$$\frac{w_r}{MPF_{si}^r} = \frac{w_k}{MPF_{si}^k} \iff \frac{w_r}{w_k} = \frac{f_r(F_{si}^1, \ldots, F_{si}^M)}{f_k(F_{si}^1, \ldots, F_{si}^M)} = \frac{f_r(1, F_{si}^2/F_{si}^1, \ldots, F_{si}^M/F_{si}^1)}{f_k(1, F_{si}^2/F_{si}^1, \ldots, F_{si}^M/F_{si}^1)}, \tag{A.7'}$$

where $w_r > 0$ denotes the rental price of factor r, MPF_{si}^r denotes the marginal product of factor r for firm i in sector s, and f_r denotes the partial derivative of f with respect to its r-th argument. The final equality uses the fact that f_r is homogeneous of degree zero, since f is homogeneous of degree one.

Setting $r = 1, 2, \ldots, k-1, k+1, \ldots, M$ in Equation (A.7') yields a system of $M-1$ equations in the $M-1$ unknowns $F_{si}^2/F_{si}^1, \ldots, F_{si}^M/F_{si}^1$. Since the cost-minimization problem admits a unique interior solution by assumption, the FOCs characterize the optimum, implying a unique solution to this system. Hence, because all firms face the same optimization problem, they choose the same relative input mix—i.e., $F_{si}^r/F_{si}^k = F_{s'j}^r/F_{s'j}^k$ for any firms i and j, sectors s and s', and factors r and k.

C.1 Derivation of the model

Given the similarity to the baseline model, we do not present the full derivation here. Instead, we solely highlight the differences relative to the baseline Cobb-Douglas case.

Households. The household side remains unchanged.

Final good producers. The original formulation remains valid, with Equations (1), (2), and (3) still holding.

Intermediate goods producers. Firms' production function is now represented by Equation (5'), replacing (5). However, sectoral output Y_s continues to be given by Equation (4). Since marginal cost remains invariant to output, Equation (6) from the Cobb-Douglas case still applies, albeit with a distinct marginal cost function. Finally, the SOC remains as in Equation (A.6) and thus holds, as before, for low values of $\eta > 0$—that is, for any $\eta \in \left(0, \frac{1+\underline{\underline{A}}/\overline{A}}{1-\underline{\underline{A}}/\overline{A}}\right]$, as shown in Proposition A.4.

Equilibrium allocation. From the result established at the beginning of this appendix, based on Equation (A.7'), we have $F_{si}^r/F_{si}^k = F_{sj}^r/F_{sj}^k$ and $MC_{si} = \frac{w_k}{MPF_{si}^k} = \frac{w_k}{A_{si} f_k(1, F_{si}^2/F_{si}^1, \ldots, F_{si}^M/F_{si}^1)}$ for any firms i and j, sector s, and factors r and k. Consequently, $\frac{MC_{si}}{MC_{sj}} = \frac{A_{sj}}{A_{si}}$, as in the baseline Cobb-Douglas case, implying that Equations (A.1), (A.2), (A.3), (7), and (8) remain valid.

Aggregate productivity and misallocation. The aggregate production function is no longer given by Equation (9). However, given Equations (4) and (5'), the new expression is

$$Y = \int_0^N A_i f(F_i^1, \ldots, F_i^M) di = \overline{A}\Omega f(F_1, \ldots, F_M), \tag{9'}$$

where $F_r \equiv \int_0^N F_i^r di$ and $\Omega \equiv \int_0^N \frac{f(F_i^1, \ldots, F_i^M)}{f(F_1, \ldots, F_M)} \left(A_i/\overline{A}\right) di$.

As established at the beginning of this appendix, cost-minimization implies that every active firm i selects the same relative input mix, so that

$$\theta_{ri} \equiv \frac{F_i^r}{\int_0^N F_j^r dj} = \frac{F_i^k \frac{F_i^r}{F_i^k}}{\int_0^N F_j^k \frac{F_j^r}{F_j^k} dj} = \frac{F_i^k \frac{F_i^r}{F_i^k}}{\int_0^N F_j^k \frac{F_i^r}{F_i^k} dj} = \frac{F_i^k}{\int_0^N F_j^k dj} \equiv \theta_{ki} \tag{A.8'}$$

for each pair of inputs r and k. This allows us to define $\theta_i \equiv \theta_{1i} = \theta_{2i} = \cdots = \theta_{Mi}$. As a result, using Equations (5') and (9'), we obtain

$$s(A_i) = \frac{Y_i}{Y} = \frac{A_i f(\theta_i F_1, \ldots, \theta_i F_M)}{\overline{A}\Omega f(F_1, \ldots, F_M)} = \frac{A_i \theta_i}{\overline{A}\Omega} \rightarrow \theta_i = \overline{A}\Omega \frac{s(A_i)}{A_i}, \tag{A.10'}$$

which is identical to Equation (A.10). As a consequence, the aggregate TFP $\overline{A}\Omega$ continues to correspond to the quantity-weighted harmonic mean of firms' productivity shown in Equation (A.11), with allocative efficiency Ω still given by Equation (10), and thus retaining the same properties as in the baseline model.

Markups. Since Equations (7) and (A.10) still hold, firm i's markup, $\mu_i = \left(1 - \frac{s_i}{\eta}\right)^{-1}$, and the cost-weighted average markup, $\mu \equiv \int_0^N \left(\frac{\sum_{k=1}^M F_i^k w_k}{\sum_{k=1}^M F_k w_k}\right) \mu_i di = \int_0^N \theta_i \mu_i di$, continue to be given by Equations (11) and (12), respectively.

Entry decision and free-entry equilibrium. Equations (13) and (A.12) remain valid if the Cobb-Douglas composite of aggregate inputs, $K^\alpha L^{1-\alpha}$, is replaced by the new composite, $f(F_1, \ldots, F_M)$. Formally,

$$\bar{\pi} = f(F_1, \ldots, F_M) \left(\frac{\overline{A}\Omega - \underline{A}}{N}\right) \tag{13'}$$

$$\check{f}_e = f(F_1, \ldots, F_M) \left(\frac{\overline{A}\check{\Omega} - \underline{\underline{A}}}{\check{N}}\right), \tag{A.12'}$$

where $\check{\Omega}$ is defined as before.[32]

C.2 Quantitative strategy for TFP decomposition

The model's solution for the target moments—aggregate TFP, $\overline{A}\Omega$, and the cost-weighted average markup, μ—is identical to that in the Cobb-Douglas case, as given by Equations (10) and (12). Consequently, the quantitative strategy for TFP decomposition described in Section 3 remains broadly valid. Specifically, given the shape of the productivity distribution, decomposing TFP requires calibrating only the productivity bounds $\underline{A}$ and $\overline{A}$, which proceeds exactly as before by applying the algorithm described in Section 3.1. Importantly, the existence and uniqueness results established for the Cobb-Douglas case in Appendix B continue to hold in this more general setting.

The only substantive difference introduced by this generalization concerns the empirical computation of the target moments. On the one hand, aggregate TFP is still identified as a residual from the aggregate production function, which is now given by Equation (9'):

$$\overline{A}\Omega = \frac{Y}{f(F_1, \ldots, F_M)}.$$

On the other hand, the cost-weighted average markup μ continues to be pinned down by the ratio of the labor elasticity of output to the labor income share. Formally, if labor is one of the inputs, then for any intermediate firm i,

$$\mu_i = \frac{Y_i p}{L_i w} \frac{L_i MPL_i}{Y_i} = \frac{\alpha_L}{LS_i},$$

[32] The derivation of Equation (13') follows the same steps as in Appendix A.7, except that Equation (9') is used in place of Equation (9).

where $\alpha_L \equiv \frac{L_i MPL_i}{Y_i}$ denotes the elasticity of output with respect to labor. As a result,

$$\mu = \int_0^N \theta_i \mu_i di = \int_0^N \left(\frac{L_i}{L}\right)\left(\alpha_L \frac{Y_i p}{L_i w}\right) di = \alpha_L \frac{p\int_0^N Y_i di}{Lw} = \frac{\alpha_L}{LS}. \tag{15'}$$

Under the generalized production function (5'), computing $\alpha_L \equiv \frac{L_i MPL_i}{Y_i}$ requires information on the relative aggregate quantities of factors, since

$$\alpha_L = \frac{L_i f_1(L_i, F_i^2, \ldots, F_i^M)}{f(L_i, F_i^2, \ldots, F_i^M)} = \frac{f_1(1, F_i^2/L_i, \ldots, F_i^M/L_i)}{f(1, F_i^2/L_i, \ldots, F_i^M/L_i)} = \frac{f_1(1, F_2/L, \ldots, F_M/L)}{f(1, F_2/L, \ldots, F_M/L)},$$

where we assume that labor corresponds to input 1 and use the fact that all firms choose the same relative factor quantities, as established at the beginning of this appendix. By contrast, in the Cobb-Douglas case, α_L is simply the exponent on labor in the production function. Indeed, if $f(L,K) = K^\alpha L^{1-\alpha}$, we have $\alpha_L = \frac{f_1(1,K/L)}{f(1,K/L)} = \frac{(1-\alpha)(K/L)^\alpha}{(K/L)^\alpha} = 1-\alpha$, which implies $\mu = \frac{1-\alpha}{LS}$, exactly as in Equation (15). Thus, unlike in the Cobb-Douglas case, α_L is not generally time invariant under production function (5'), even if the parameters of the technology remain constant. As a consequence, growth in the average markup μ need not mirror growth in the labor share.

C.3 Calibration of the production function

In the Cobb-Douglas case, cost minimization implies that α equals capital's share in total factor costs, that is, $\alpha = \frac{Kr}{Kr+Lw}$ (Equation (A.9)). This property allows the production function to be calibrated using cost data, as discussed in Section 3.3.[33] A similar approach can be employed under the generalized production function (5').

Formally, since $\mu_i = \alpha_L / LS_i$, it follows that

$$\alpha_L = \frac{L_i w}{Y_i p/\mu_i} = \frac{\theta_i Lw}{\sum_{k=1}^M \theta_i F_k w_k} = \frac{Lw}{\sum_{k=1}^M F_k w_k}, \tag{A.9'}$$

where we use the fact that $F_i^k = \theta_i F_k$ for any input k and firm i.[34] As a result, the production function can be calibrated by choosing the parameters of f so that $\alpha_L = \frac{f_1(1,F_2/L,\ldots,F_M/L)}{f(1,F_2/L,\ldots,F_M/L)}$ matches the observed labor cost share. Cost share data for other inputs can also be incorporated into the calibration by using $\alpha_r = \frac{F_r w_r}{\sum_{k=1}^M F_k w_k}$ for all factors r. However, only $M-1$ cost shares are informative, since by definition $\sum_{r=1}^M \alpha_r = 1$.

[33] In models with input-output linkages, aggregate production elasticities generally depend on the production network and cannot be recovered from aggregate factor cost shares alone (Baqaee and Farhi, 2020). In this appendix, we abstract from this complexity and strictly follow the simple structure of our model.

[34] As expected, Equation (A.9') is consistent with Equation (A.9): in the baseline case, $1-\alpha = \alpha_L = \frac{Lw}{Kr+Lw}$.

To illustrate this calibration procedure, consider the CES production function

$$Y_i = A_i \left[\alpha K_i^{\frac{\sigma-1}{\sigma}} + (1-\alpha) L_i^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{C.1}$$

where $\sigma > 0$ denotes the elasticity of substitution and $\alpha \in (0,1)$. In this case, the elasticity of output with respect to labor is given by

$$\alpha_L = \frac{f_1(1, K/L)}{f(1, K/L)} = \frac{\left[\alpha (K/L)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\right]^{\frac{1}{\sigma-1}} (1-\alpha)}{\left[\alpha (K/L)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\right]^{\frac{\sigma}{\sigma-1}}} = \frac{1-\alpha}{\alpha (K/L)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)}. \tag{C.2}$$

As expected, $\alpha_L \to 1-\alpha$ as $\sigma \to 1$, since the CES production function (C.1) converges to the baseline Cobb-Douglas specification (5) in this limit. Hence, given data on K and L and values for the parameters α and σ, Equation (C.2) allows α_L to be computed directly, making it possible to calibrate α and σ by matching α_L to labor cost share data.

D Bounds for the capital elasticity

In a setting without input-output linkages—as in our model—cost minimization implies (up to a first-order approximation) that $\alpha \approx \frac{Kr}{Kr+Lw}$.[35] With input-output linkages, however, aggregate production elasticities generally depend on the production network and cannot be recovered from aggregate factor cost shares alone (Baqaee and Farhi, 2020). In either case, calibrating α requires measuring capital costs, Kr, which is challenging because national accounts do not separately report capital expenses and economic profits.

Because estimating capital expenses is notoriously difficult, Vollrath (2024) proposes plausible bounds for α in the US, based on extreme assumptions about capital costs. The upper bound is obtained under the no-profit assumption—so that all value added not paid as labor compensation accrues to capital—implying $\alpha = 1 - LS$. The lower bound assumes instead that capital costs account only for depreciation. In his baseline analysis, Vollrath (2024) computes capital expenses under these extreme assumptions at the industry level and then uses input-output information to recover aggregate production elasticities, following the framework of Baqaee and Farhi (2020). He also computes the bounds directly from aggregate cost shares. The no-profit upper bound necessarily coincides with $1 - LS$ (Baqaee and Farhi, 2020). This equivalence does not hold under alternative assumptions about capital costs, but Vollrath (2024) finds that the depreciation-based lower bound computed from aggregate cost-share data is only slightly lower—by about 0.03 on average.

Motivated by these findings, we use PWT 11.0 to construct analogous bounds for the global economy based on aggregate cost shares.[36] Throughout the paper, we focus on non-oil countries and therefore exclude oil producers from these bounds.[37] Figure D.1 reports these global bounds and compares them to the baseline US bounds in Vollrath (2024). The lower bound—constructed under the depreciation-only assumption—is broadly similar across the two cases, particularly after the 1980s. In our global estimates, it remains close to 0.20 until the early 2000s and then gradually increases to about 0.25–0.27 by the 2010s. The upper bound—obtained under the no-profit assumption—is higher for the global economy: it fluctuates around 0.35–0.38 in the 1950s and 1960s and rises steadily thereafter, reaching roughly 0.44–0.45 by the 2010s. Overall, both bounds display a clear upward trend over time, consistent with the patterns emphasized by Vollrath (2024).

These results indicate that our baseline choice $\alpha = 0.30$ is plausible, as it lies comfortably within the global bounds and is broadly consistent with the US bounds, with only minor and

[35] In our Cobb-Douglas model, this identity holds exactly, as shown in Appendix A.4.

[36] The global bounds are computed as follows. For each country-year, we construct alternative capital cost shares $KS \in \{1 - LS,\ \delta/Y\}$, where δ/Y denotes depreciation as a share of GDP. Labor shares LS are measured using the PWT *labsh* variable, while GDP (in current PPPs) Y is measured using *cgdpo*. Depreciation δ is computed as the sum of *Dc_Struc*, *Dc_Mach*, *Dc_TraEq*, and *Dc_Other* from the PWT capital detail dataset, expressed in current PPP terms. Each KS is then mapped into an implied capital elasticity as $\alpha = KS/(KS + LS)$. Global bounds are reported as year-by-year cross-country averages weighted by total factor costs $(KS + LS)Y$.

[37] As we discuss in Section 4, a country is classified as an oil producer in a given year if its oil rents exceed 10 percent of GDP.

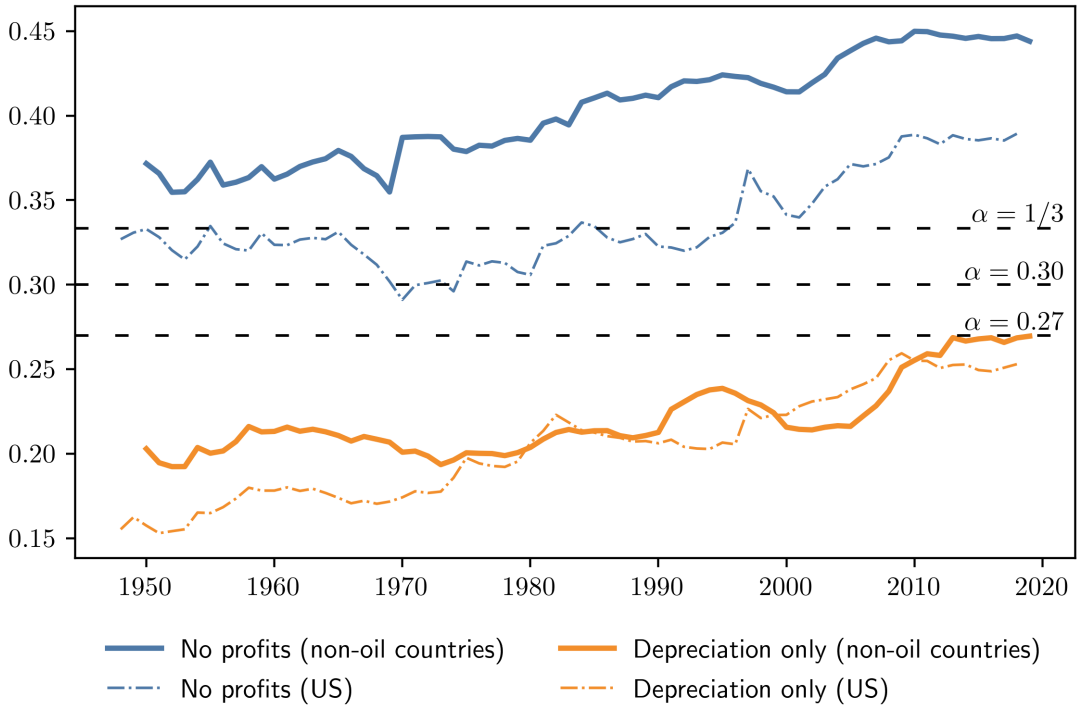


Figure D.1: Bounds for the aggregate capital elasticity, α. *Notes*: bounds for non-oil countries are based on aggregate cost shares from PWT 11.0 (1950–2019), while the US bounds correspond to the baseline estimates reported by Vollrath (2024) (1948–2018).

short-lived deviations from the upper bound in the 1970s (Figure D.1). As a robustness check, Section 8 examines two alternative calibrations for the capital elasticity, $\alpha = 1/3$ and $\alpha = 0.27$. As shown in Figure D.1, the bounds indicate that these alternative values of α correspond to relatively extreme cases within the plausible parameter space. On the upper end, while $\alpha = 1/3$ is among the most common assumptions in economics (Vollrath, 2024), it lies above the no-profits upper bound for the US over most of the sample period, falling below this bound only from the early 2000s onward. On the lower end, $\alpha = 0.27$ is essentially pinned to the global depreciation-only lower bound in recent years. Taken together, these alternative values provide a stringent robustness check, spanning a wide range of economically plausible elasticities.

E Constructing a broad-coverage labor income share measure for growth analyses

In level analyses, we rely on the baseline PWT labor share series (*labsh*) in the dataset that maximizes coverage. For growth analyses, however, this baseline series is unsuitable, as it may not accurately measure labor share growth rates, potentially distorting estimates of changes in allocative efficiency over time. This concern arises for two reasons.

First, the treatment of self-employed income underlying the baseline series is not necessarily consistent over time. The PWT considers five distinct labor share measures, available in the PWT labor detail database, which differ in their treatment of the income of self-employed workers (also referred to as mixed income):

1. Naive share (*comp_sh*): $\frac{\text{labor compensation of employees}}{\text{GDP at basic prices}}$.
2. Adj. 1, mixed income (*lab_sh1*): $\frac{\text{labor compensation of employees+mixed income}}{\text{GDP at basic prices}}$.
3. Adj. 2, partial mixed income (*lab_sh2*): $\frac{\text{labor compensation of employees}}{\text{GDP at basic prices}-\text{mixed income}}$.
4. Adj. 3, average wage(*lab_sh3*): $\frac{\left(\frac{\text{labor compensation of employees}}{\text{number of employees}}\right)(\text{number of employees+number of self-employed})}{\text{GDP at basic prices}}$.
5. Adj. 4, agriculture (*lab_sh4*): $\frac{\text{labor compensation of employees+value added in agriculture}}{\text{GDP at basic prices}}$.

The baseline labor share series is constructed by applying a set of selection rules to choose among these measures on a country-by-country and year-by-year basis. As a result, the underlying labor share concept may change over time within a given country.

Second, all five labor share measures include imputed observations to ensure coverage. Specifically, labor shares are assumed to remain constant at the beginning or end of the sample when data are unavailable, and missing observations within the sample are filled in by linear interpolation.

Against this backdrop, we construct an alternative labor share series that addresses these limitations and is therefore suitable for growth analyses. The series draws on the PWT variables *lab_sh2*, *comp_sh*, and *labsh*, excluding all imputed observations, as identified in the PWT labor detail database. The construction proceeds as follows. For each country, the growth rate of the constructed series follows the growth of the preferred PWT adjustment 2 measure (*lab_sh2*). When this measure is unavailable, we instead rely on the growth of the naive share (*comp_sh*). Once country-specific growth rates are determined, the level of the series is pinned down by matching its country-specific mean to that of the baseline PWT labor share series (*labsh*), over the longest coincident sample available.

This constructed series achieves broader sample coverage than the preferred PWT adjustment 2 measure, as it does not necessarily require data on self-employed income, which are available for a more limited set of countries. Moreover, insofar as the baseline PWT series

provides an accurate measure of the overall labor share level, anchoring the constructed series to this benchmark ensures a consistent level, even though the new measure relies in part on the naive share—which is known to underestimate the labor share by assigning no self-employed income to labor (Gollin, 2002). Crucially, the constructed series accurately captures labor share growth.

This is shown in Figure E.1, which compares annual labor share growth rates, $\Delta \ln LS$, from the preferred adjustment 2 measure with those from each of the other four labor share measures (all excluding imputations), using only country-year observations for which all five labor share growth measures are simultaneously available. Each panel also reports the 45-degree line and the Ordinary Least Squares (OLS) linear trend, along with the associated centered R^2. As can be seen, the naive share exhibits the strongest linear relationship with adjustment 2 growth, with an estimated slope close to unity. This evidence suggests that the naive share provides a reliable proxy for labor share growth when the preferred adjustment 2 measure is unavailable.

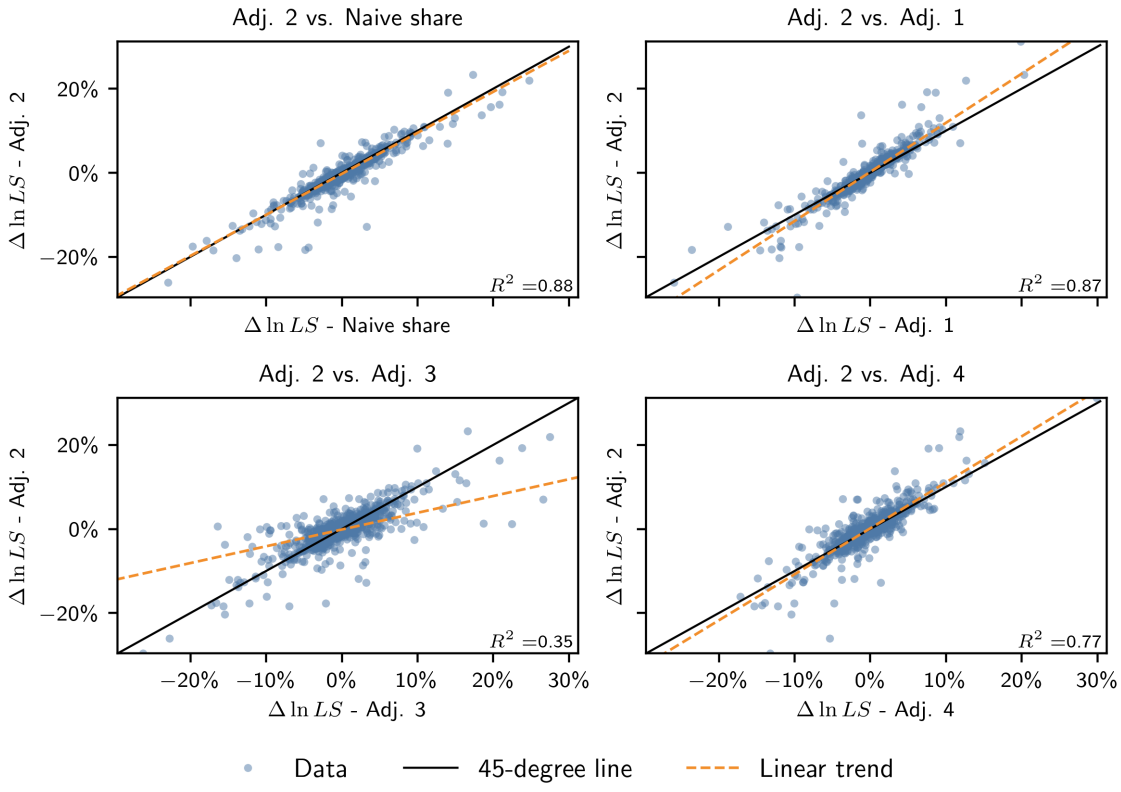


Figure E.1: Comparing alternative measures of annual labor share growth.

F Exploring the datasets for level analyses

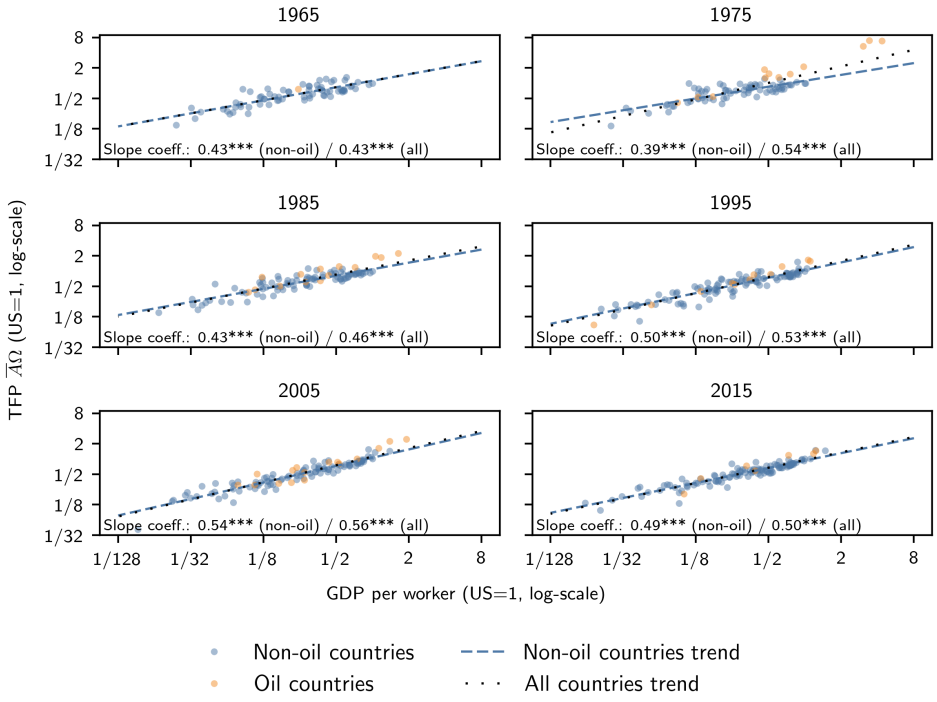


Figure F.1: TFP vs. GDP per worker – level dataset with broad coverage ($k = -1$).

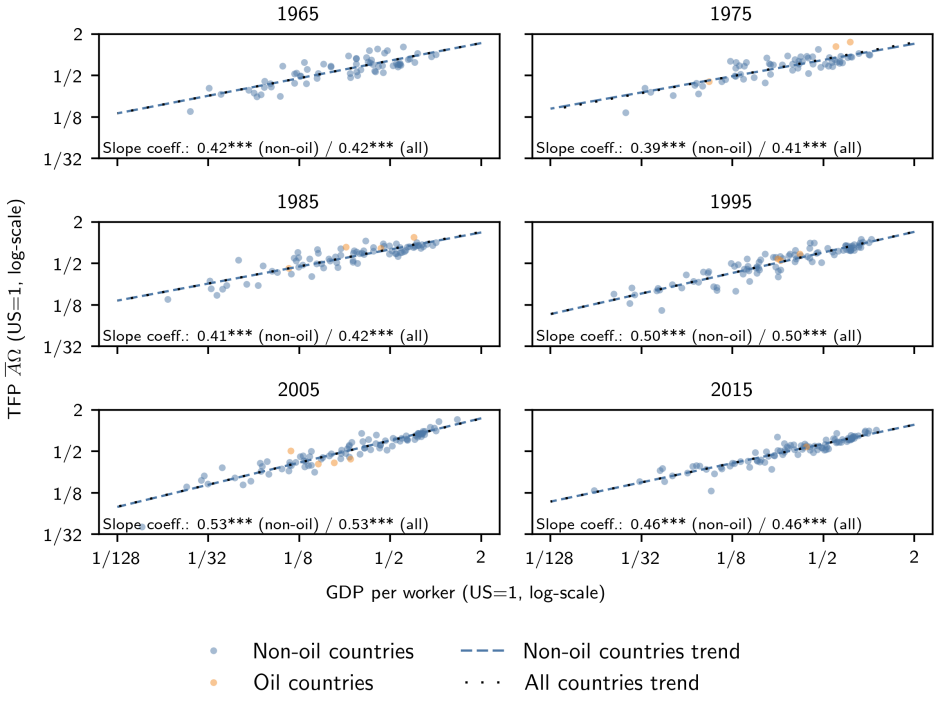


Figure F.2: TFP vs. GDP per worker – level dataset with broad coverage ($k = 3$).

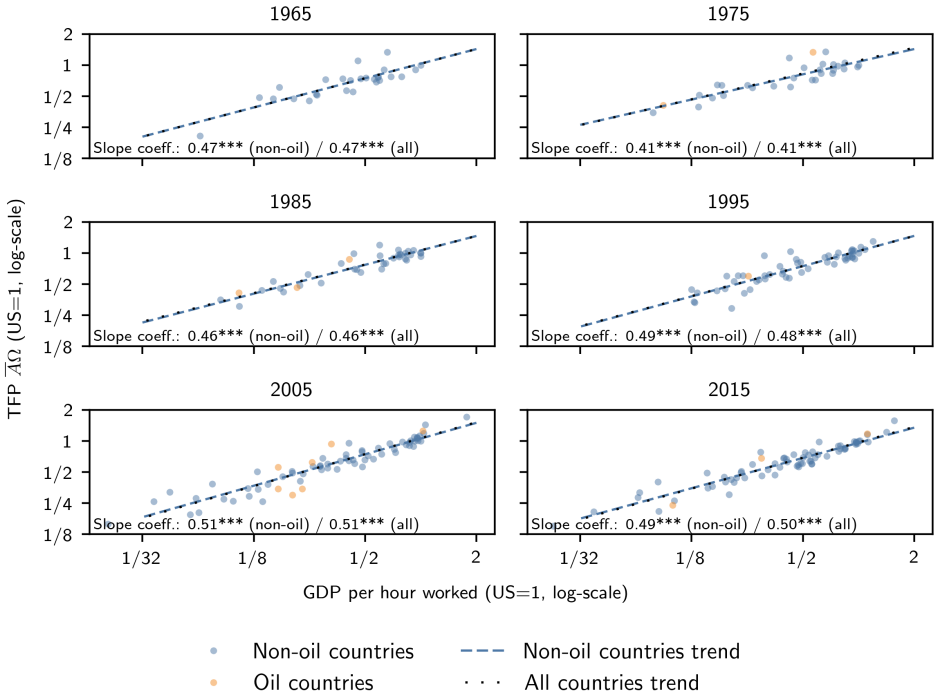


Figure F.3: TFP vs. GDP per hour worked – level dataset with best proxies ($k = -1$).

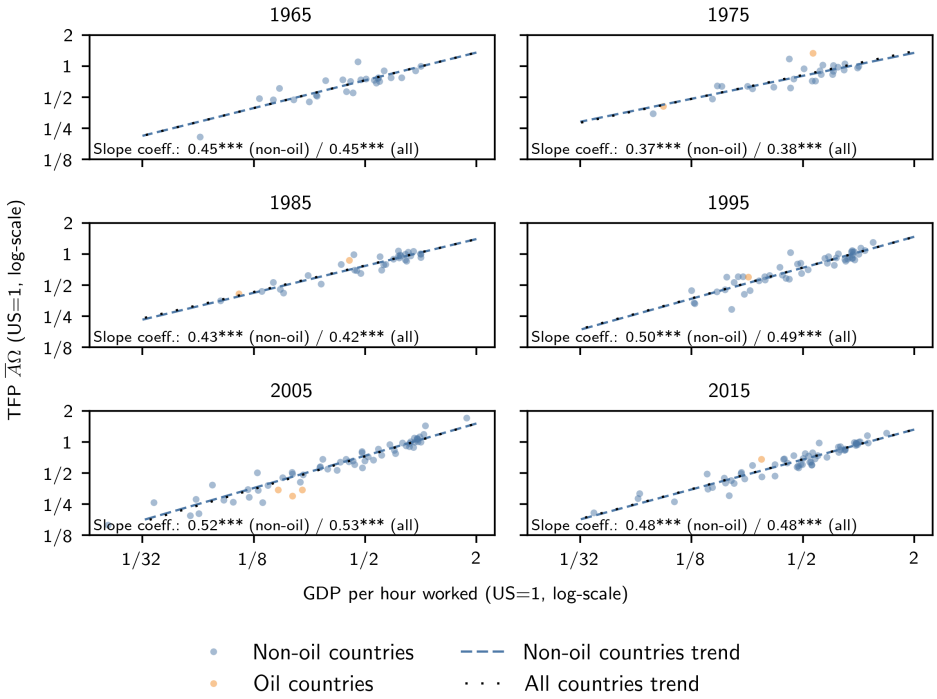


Figure F.4: TFP vs. GDP per hour worked – level dataset with best proxies ($k = 3$).

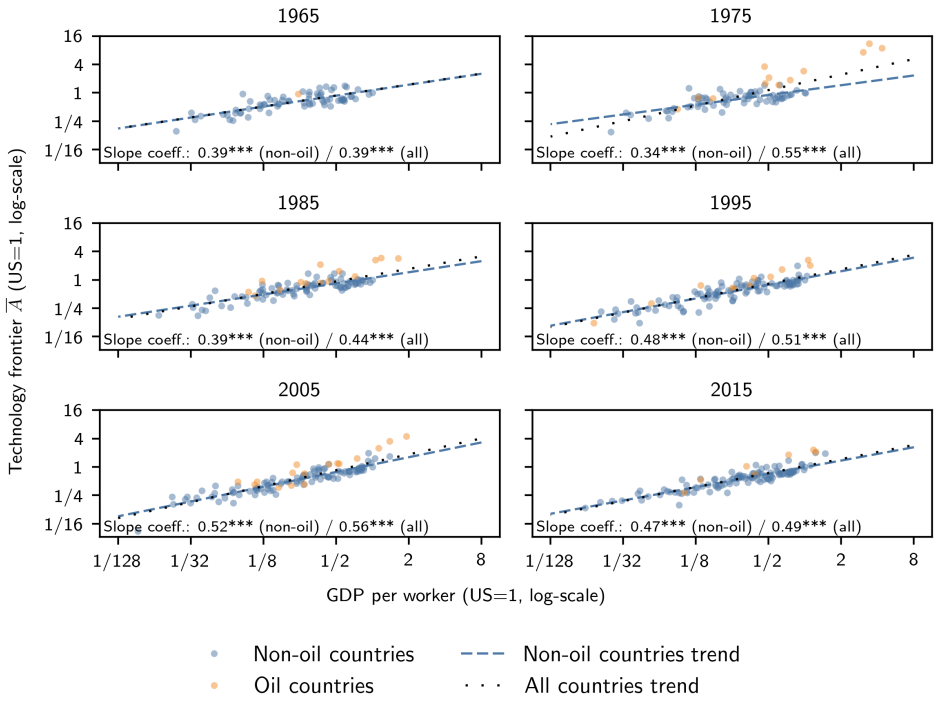


Figure F.5: Technology frontier vs. GDP per worker – level dataset with broad coverage ($k = -1$).

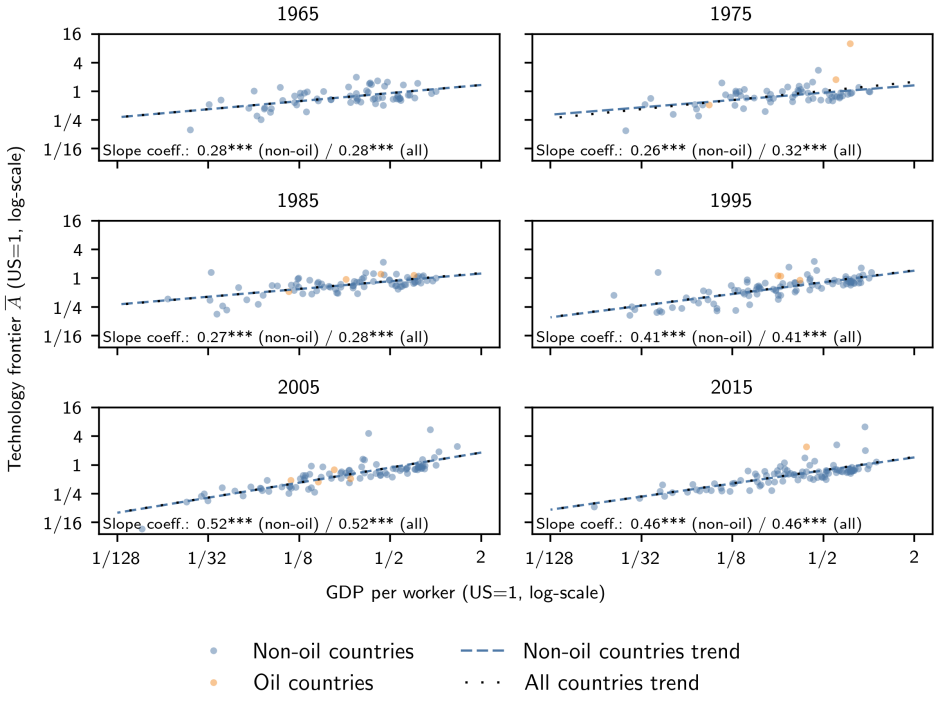


Figure F.6: Technology frontier vs. GDP per worker – level dataset with broad coverage ($k = 3$).

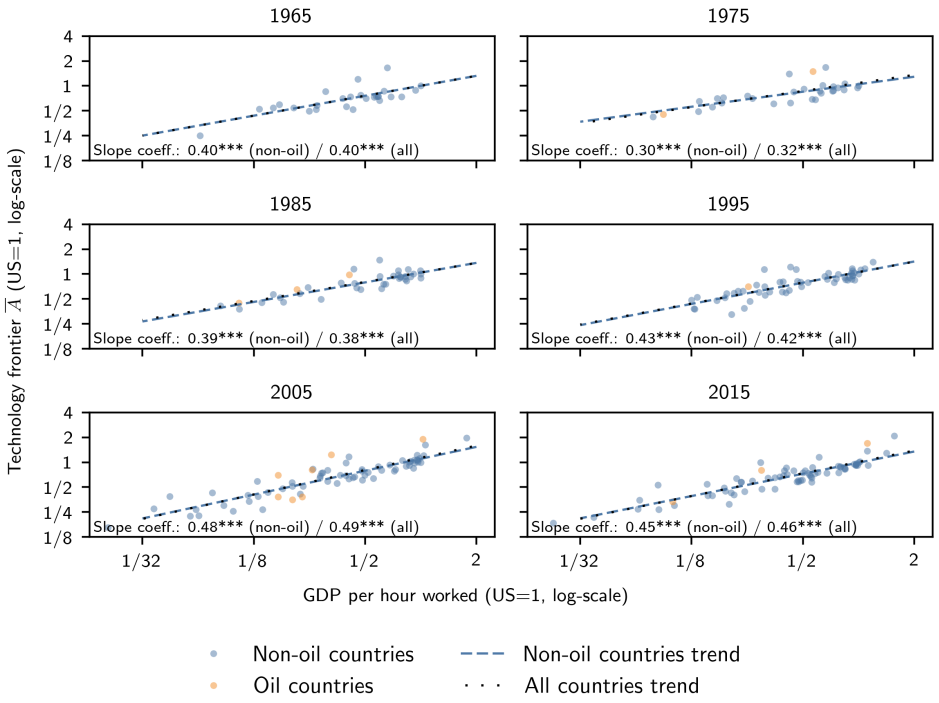


Figure F.7: Technology frontier vs. GDP per hour worked – level dataset with best proxies ($k = -1$).

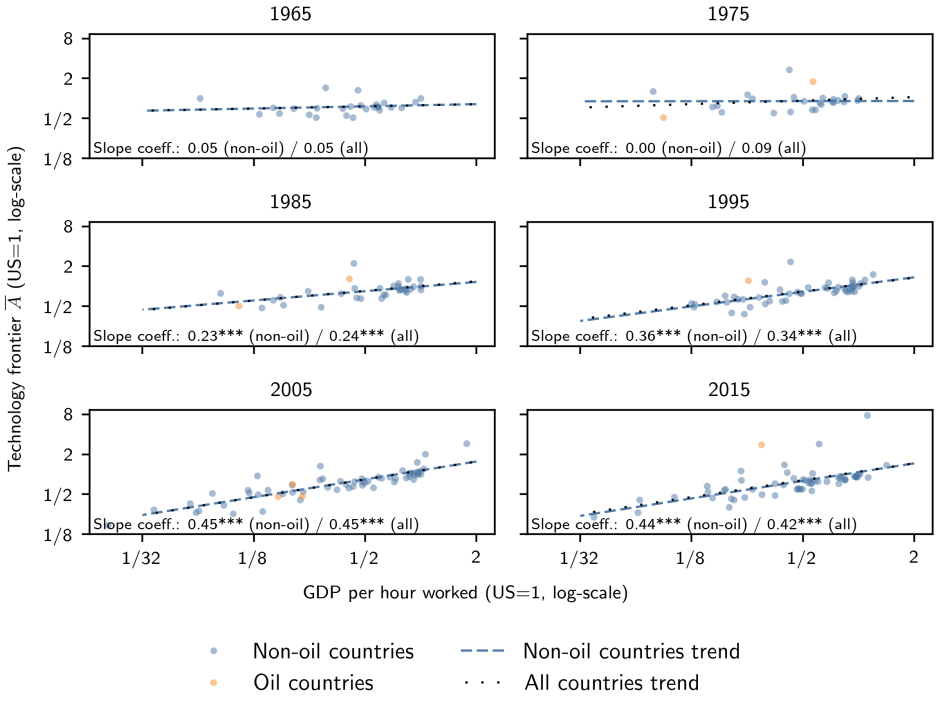


Figure F.8: Technology frontier vs. GDP per hour worked – level dataset with best proxies ($k = 3$).

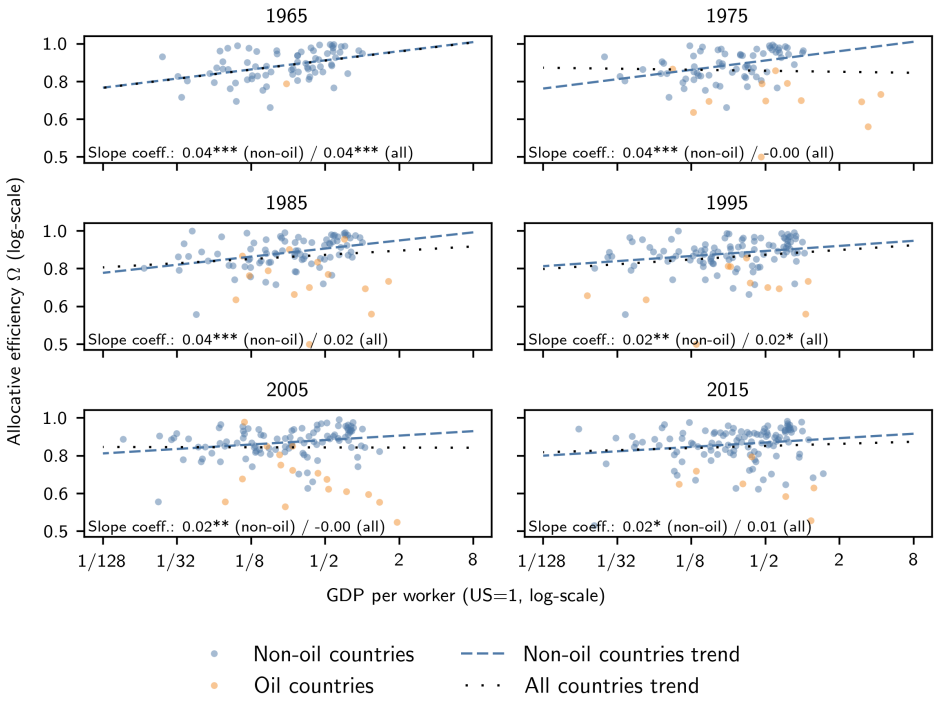


Figure F.9: Allocative efficiency vs. GDP per worker – level dataset with broad coverage ($k = -1$).

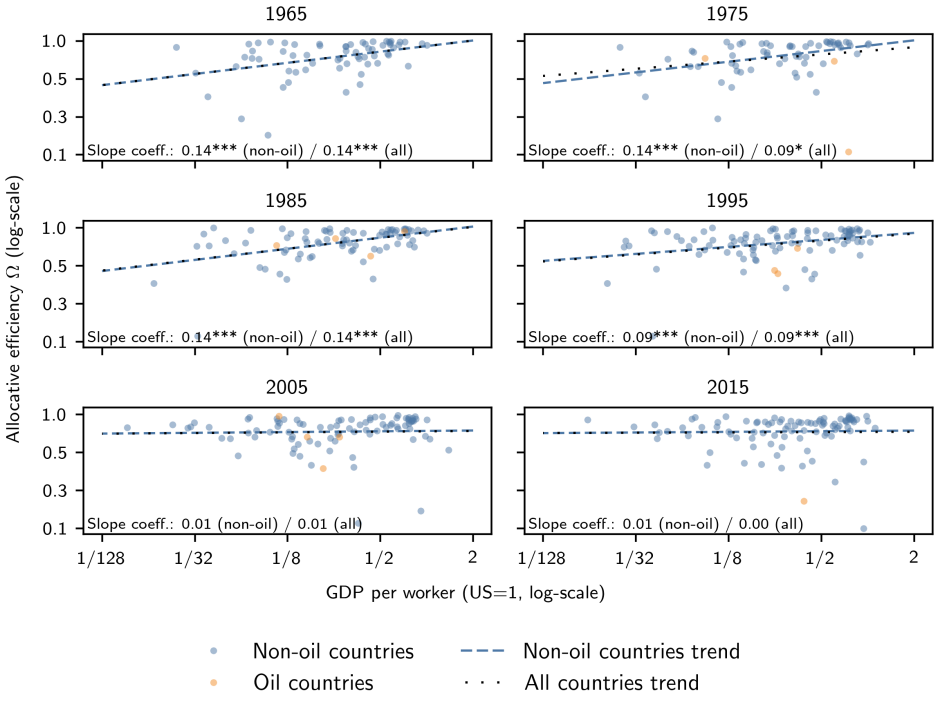


Figure F.10: Allocative efficiency vs. GDP per worker – level dataset with broad coverage ($k = 3$).

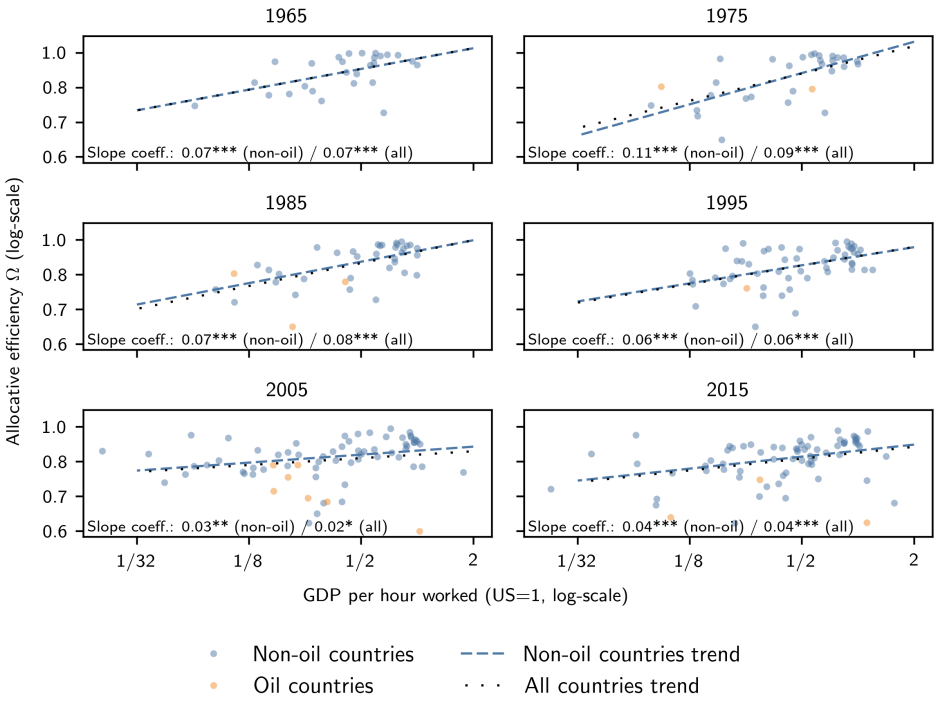


Figure F.11: Allocative efficiency vs. GDP per hour worked – level dataset with best proxies ($k = -1$).

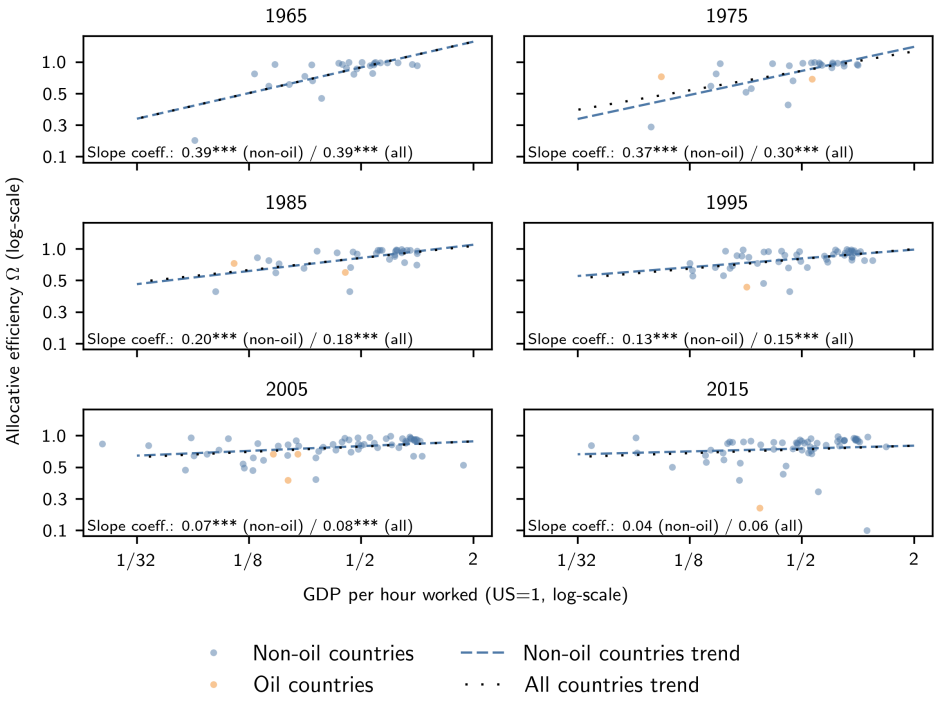


Figure F.12: Allocative efficiency vs. GDP per hour worked – level dataset with best proxies ($k = 3$).

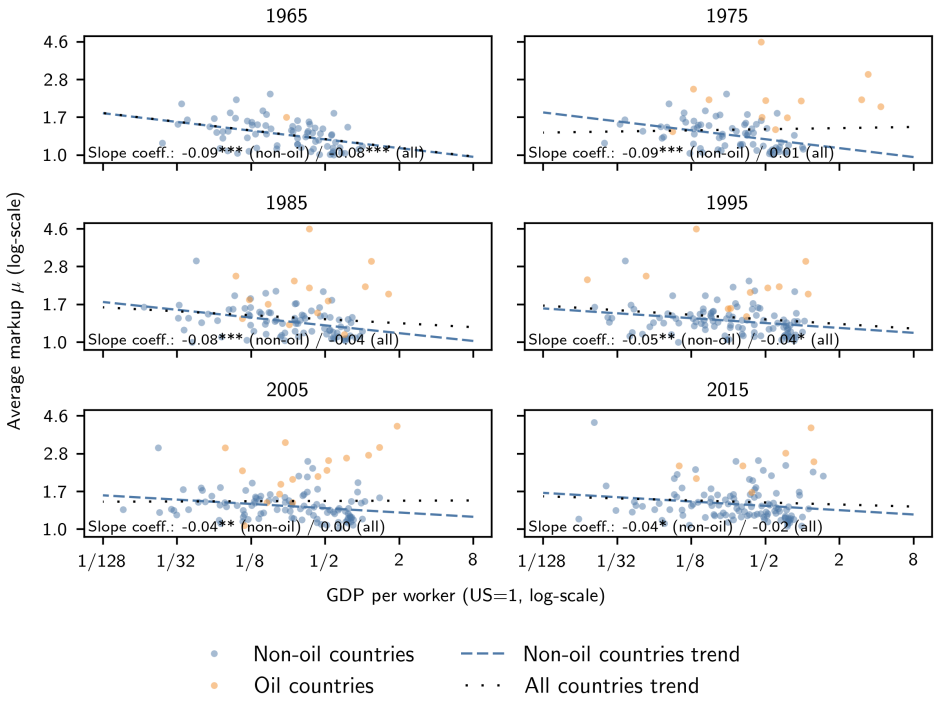


Figure F.13: Cost-weighted average markup vs. GDP per worker – level dataset with broad coverage ($k = -1$).

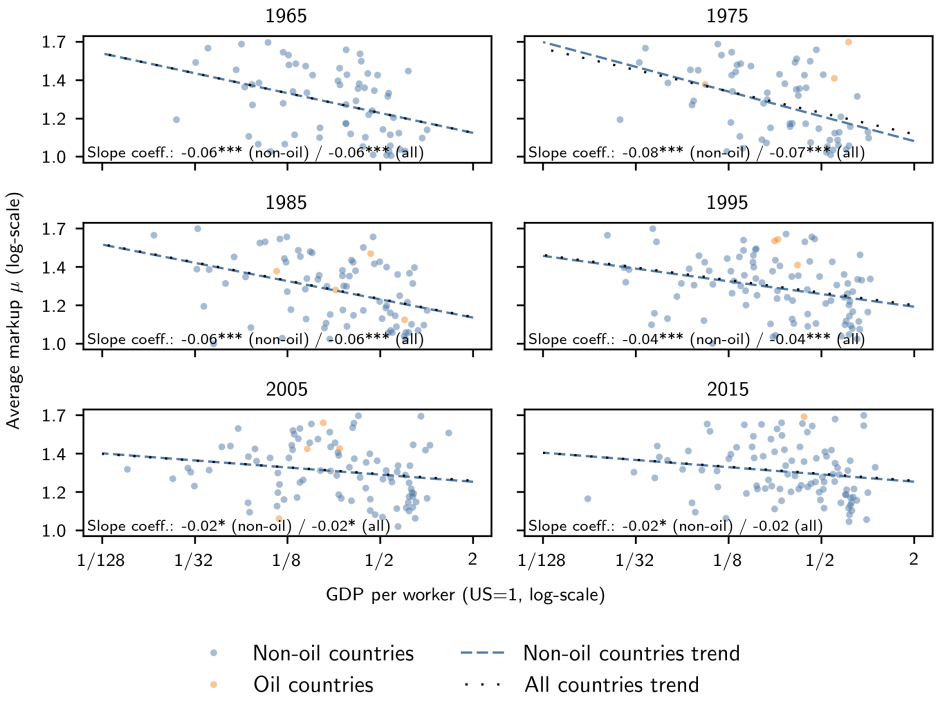


Figure F.14: Cost-weighted average markup vs. GDP per worker – level dataset with broad coverage ($k = 3$).

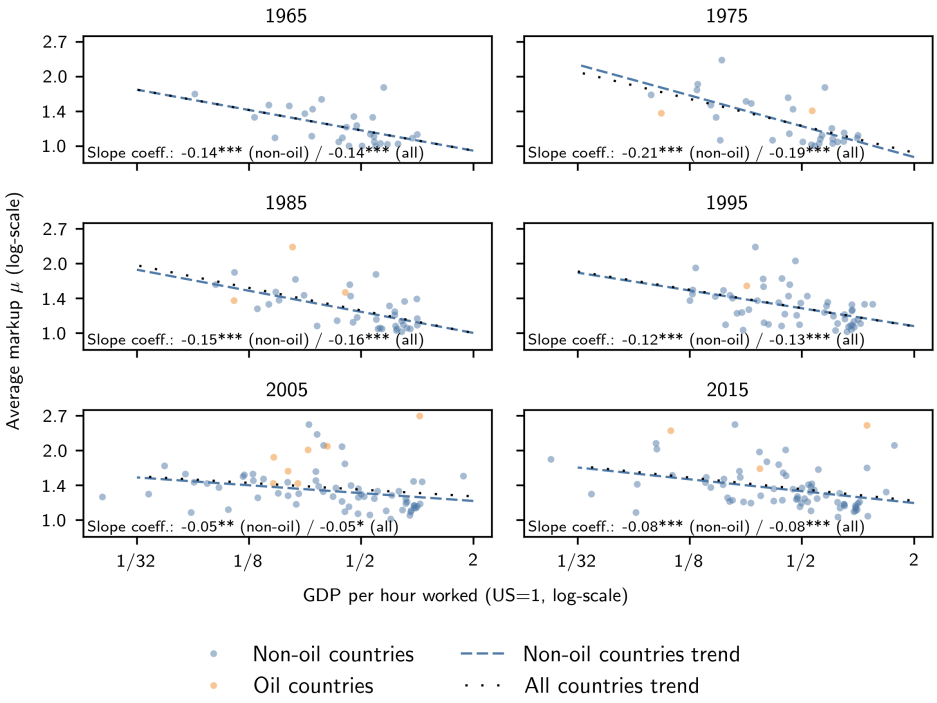


Figure F.15: Cost-weighted average markup vs. GDP per hour worked – level dataset with best proxies ($k = -1$).

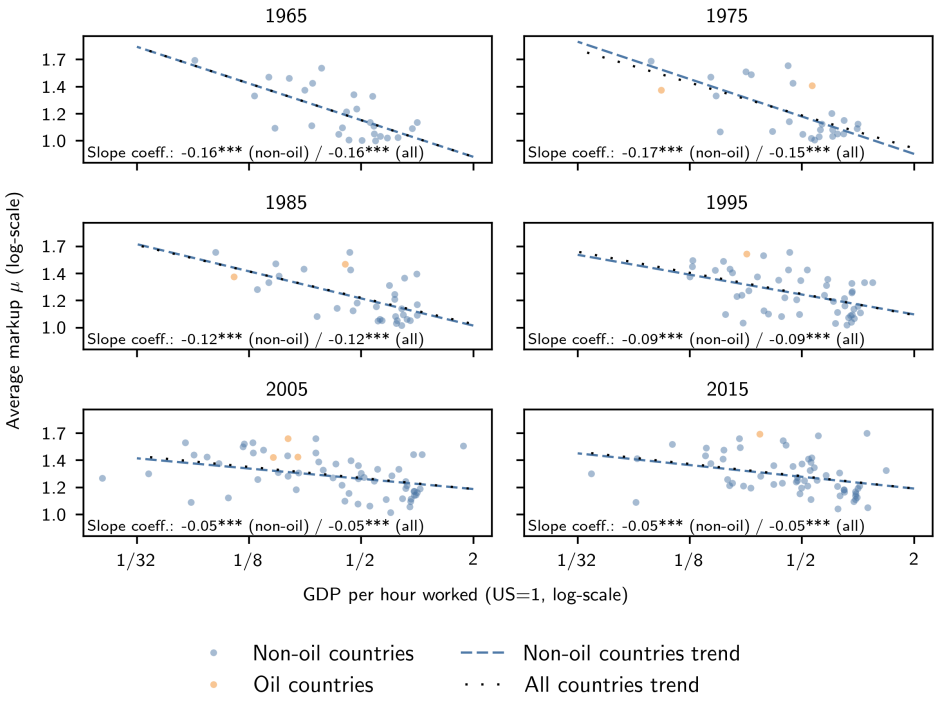


Figure F.16: Cost-weighted average markup vs. GDP per hour worked – level dataset with best proxies ($k = 3$).

G Exploring the datasets for growth analyses

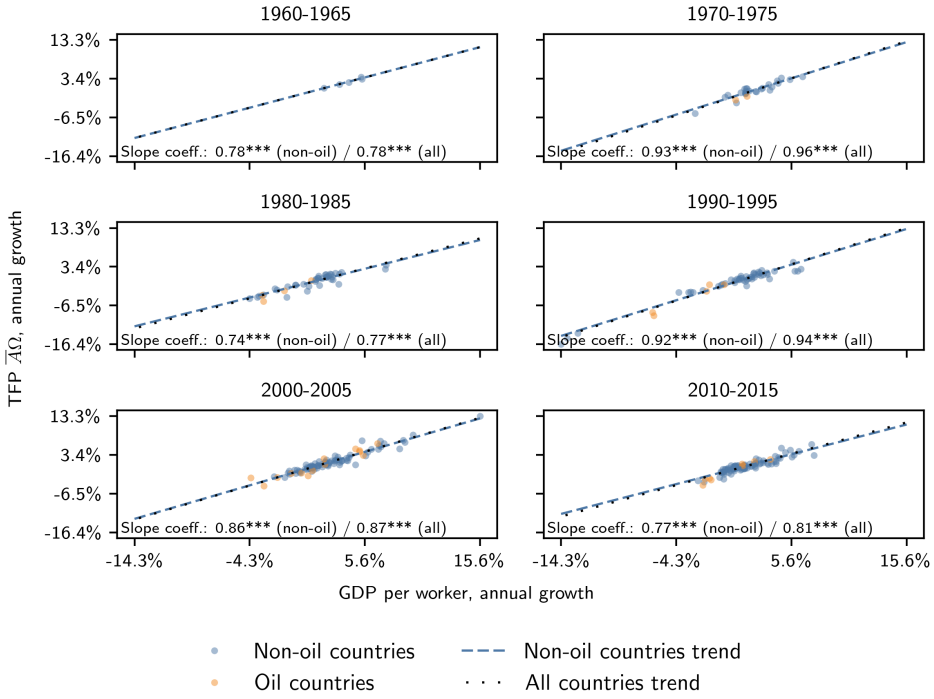


Figure G.1: TFP growth vs. GDP per worker growth – growth dataset with broad coverage ($k = -1$).

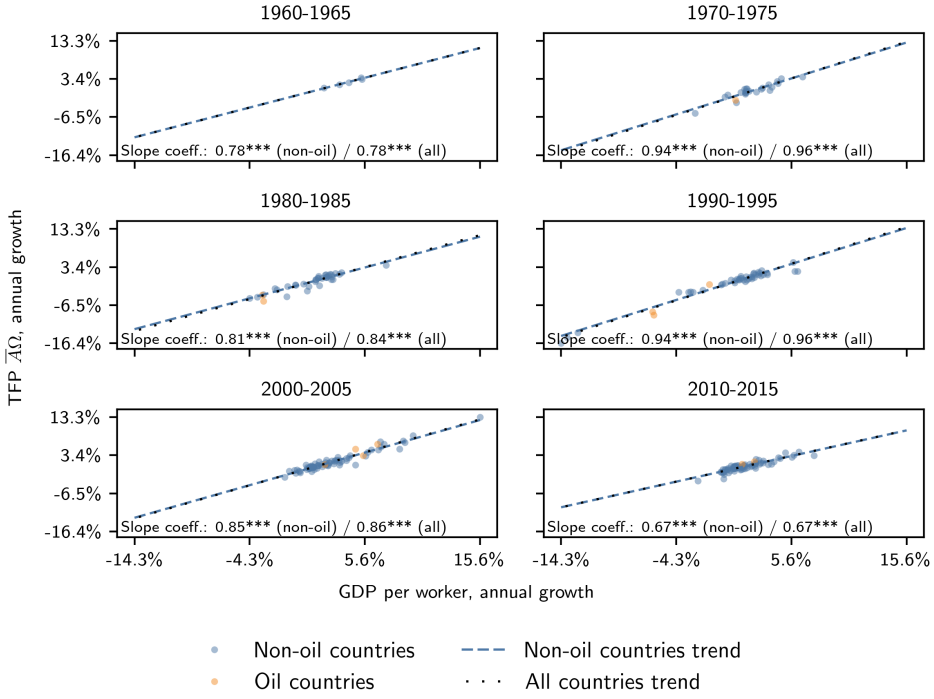


Figure G.2: TFP growth vs. GDP per worker growth – growth dataset with broad coverage ($k = 3$).

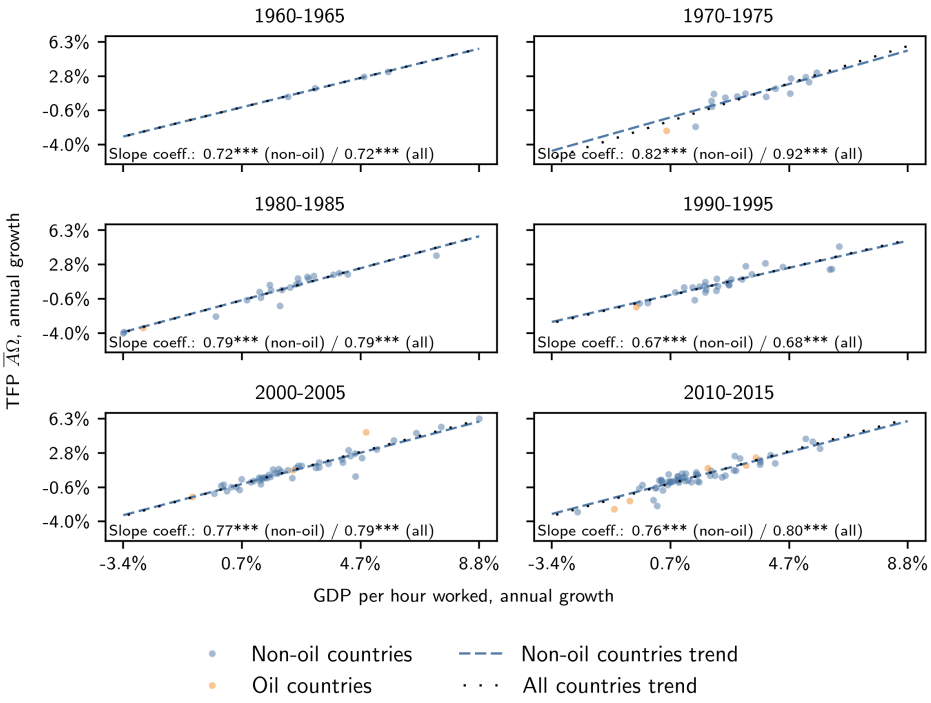


Figure G.3: TFP growth vs. GDP per hour worked growth – growth dataset with best proxies ($k = -1$).

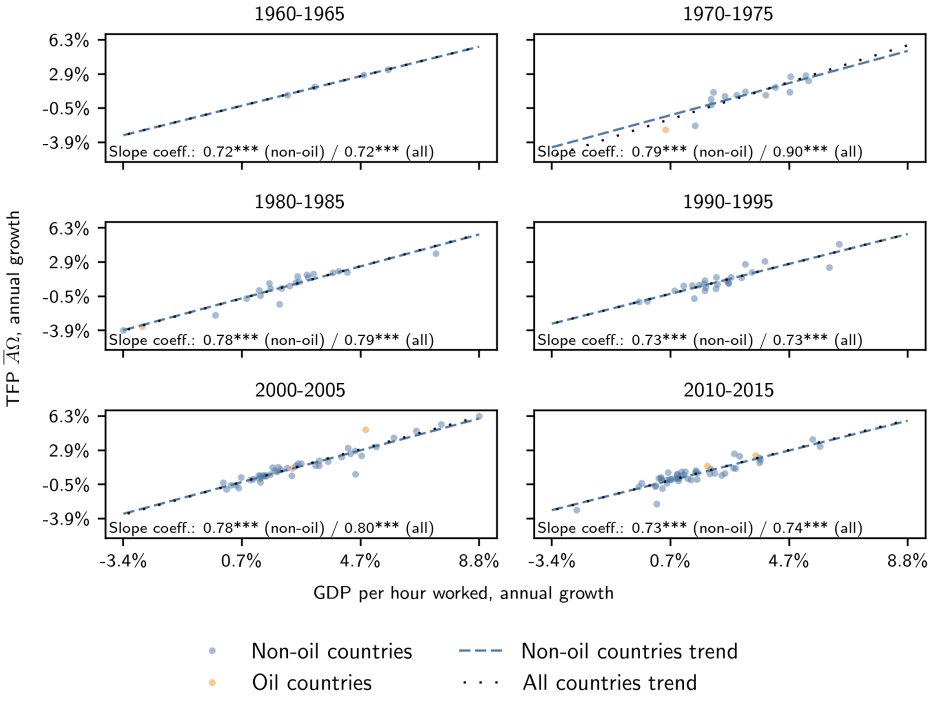


Figure G.4: TFP growth vs. GDP per hour worked growth – growth dataset with best proxies ($k = 3$).

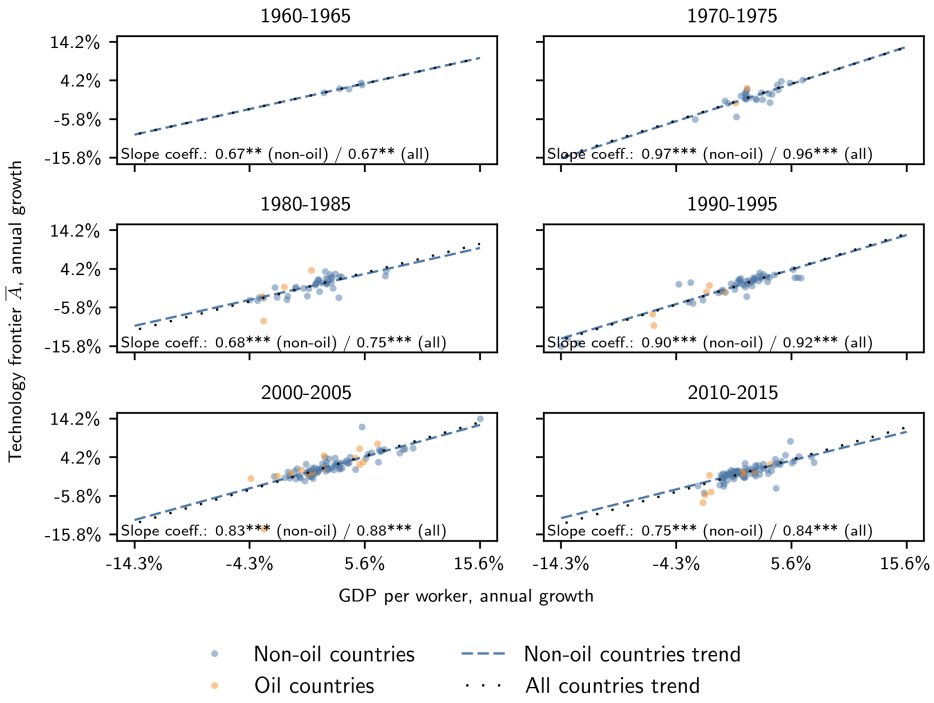


Figure G.5: Technology frontier growth vs. GDP per worker growth – growth dataset with broad coverage ($k = -1$).

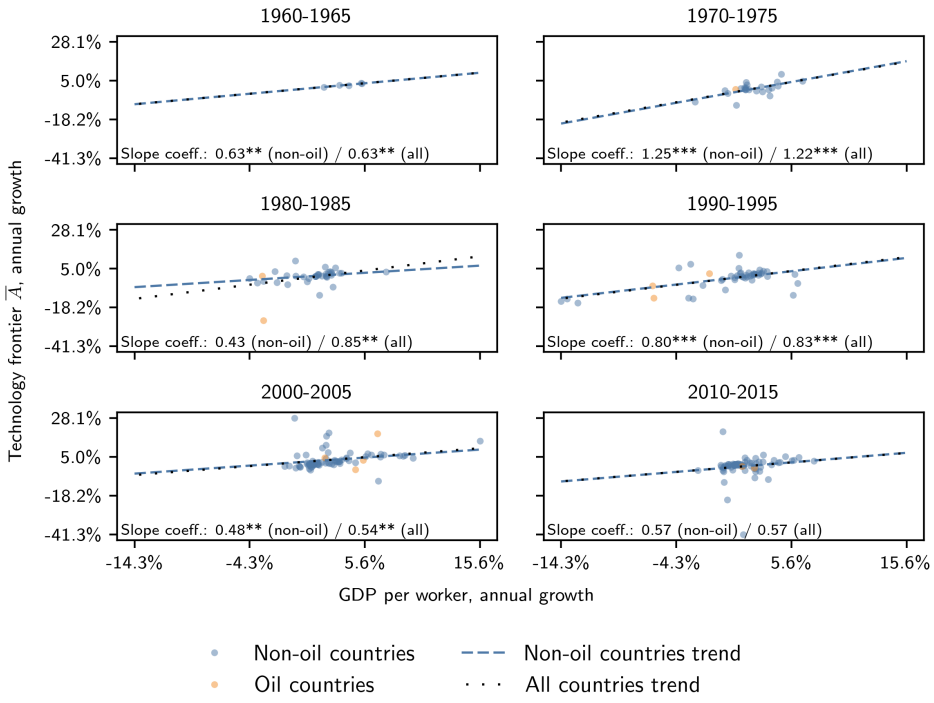


Figure G.6: Technology frontier growth vs. GDP per worker growth – growth dataset with broad coverage ($k = 3$).

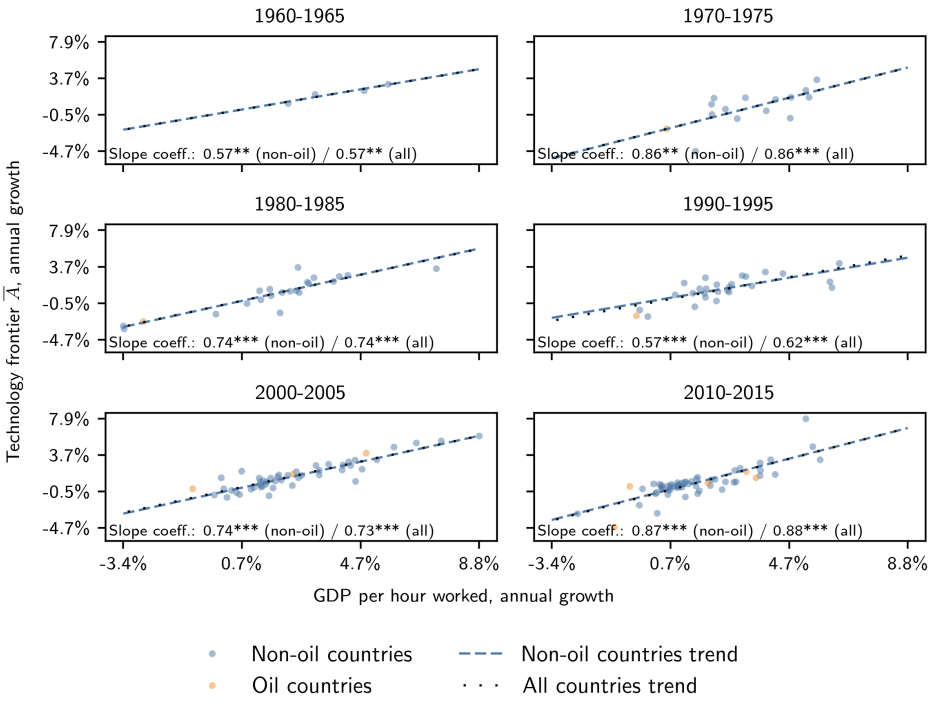


Figure G.7: Technology frontier growth vs. GDP per hour worked growth – growth dataset with best proxies ($k = -1$).

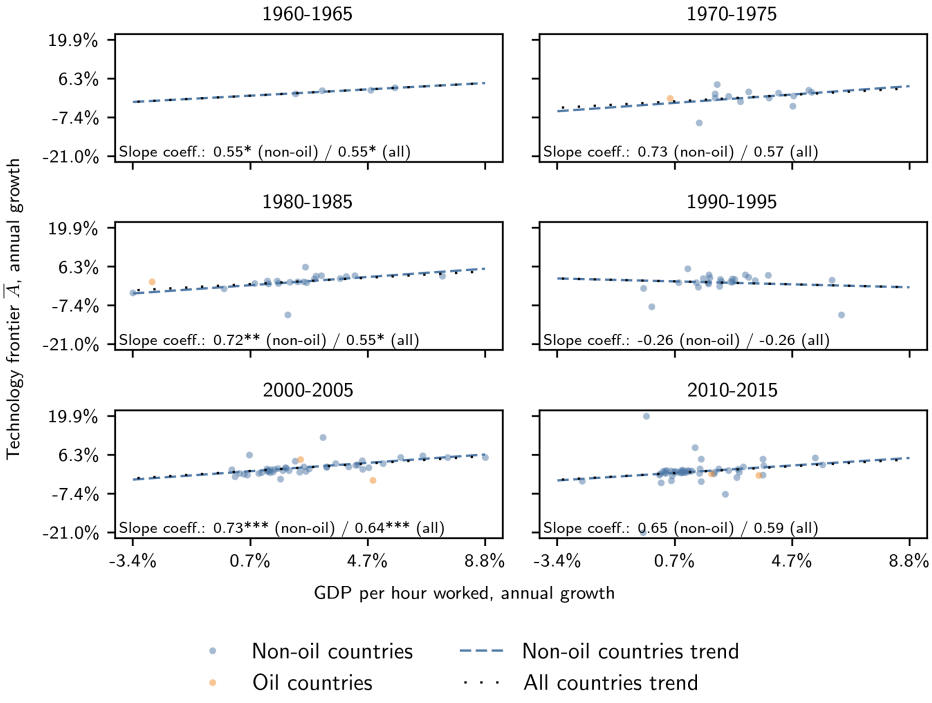


Figure G.8: Technology frontier growth vs. GDP per hour worked growth – growth dataset with best proxies ($k = 3$).

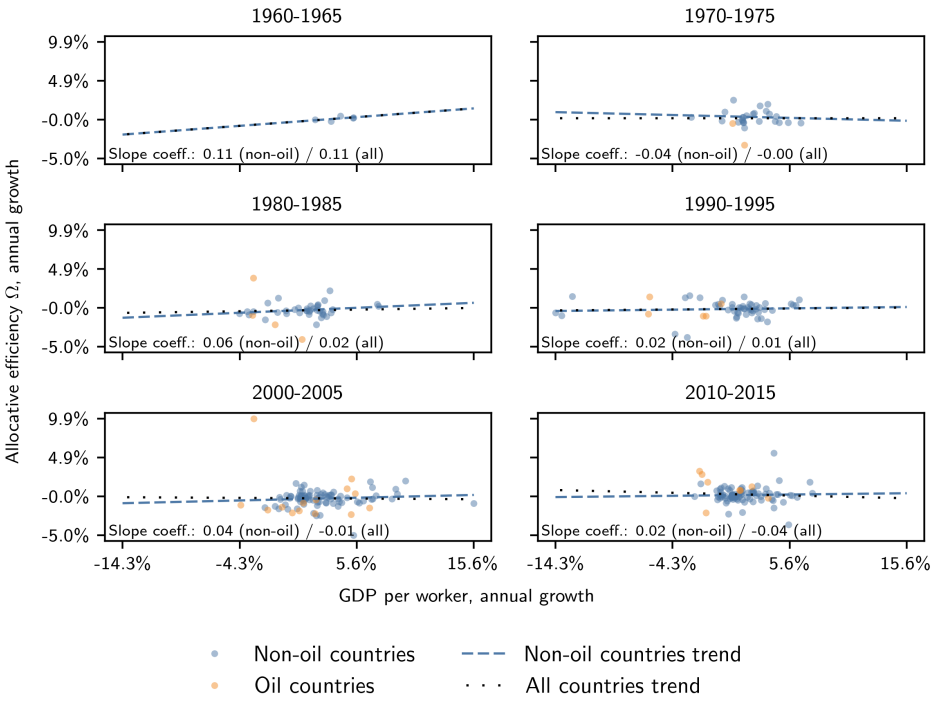


Figure G.9: Allocative efficiency growth vs. GDP per worker growth – growth dataset with broad coverage ($k = -1$).

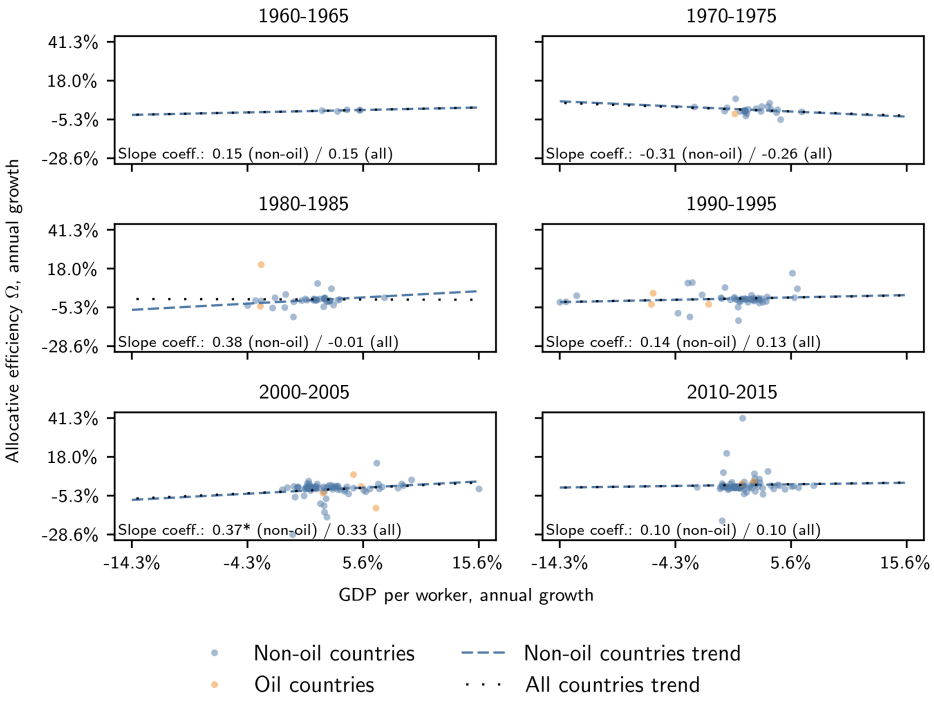


Figure G.10: Allocative efficiency growth vs. GDP per worker growth – growth dataset with broad coverage ($k = 3$).

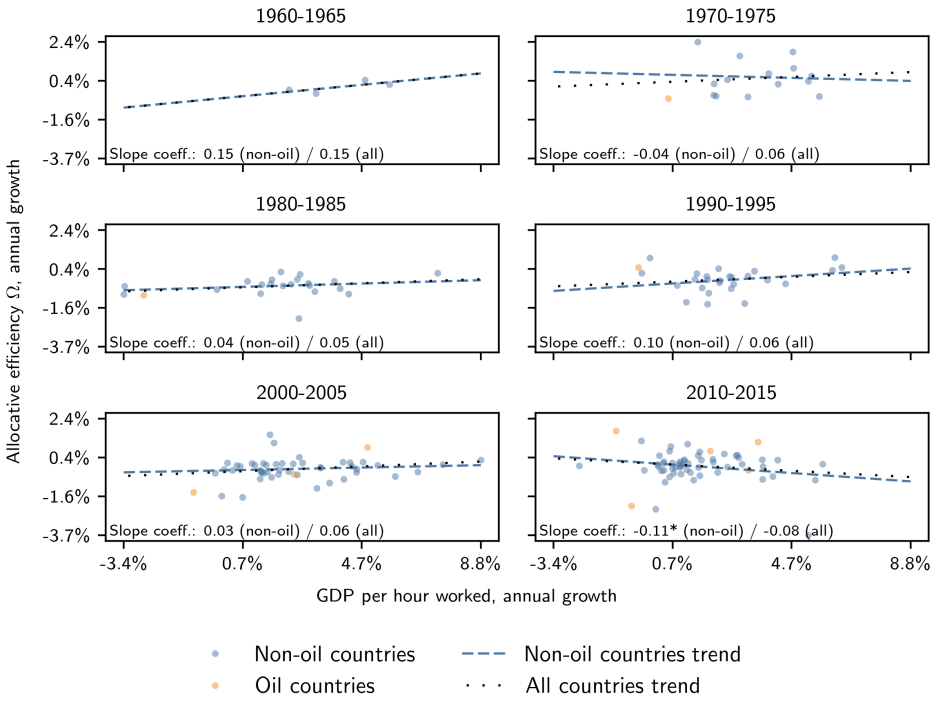


Figure G.11: Allocative efficiency growth vs. GDP per hour worked growth – growth dataset with best proxies ($k = -1$).

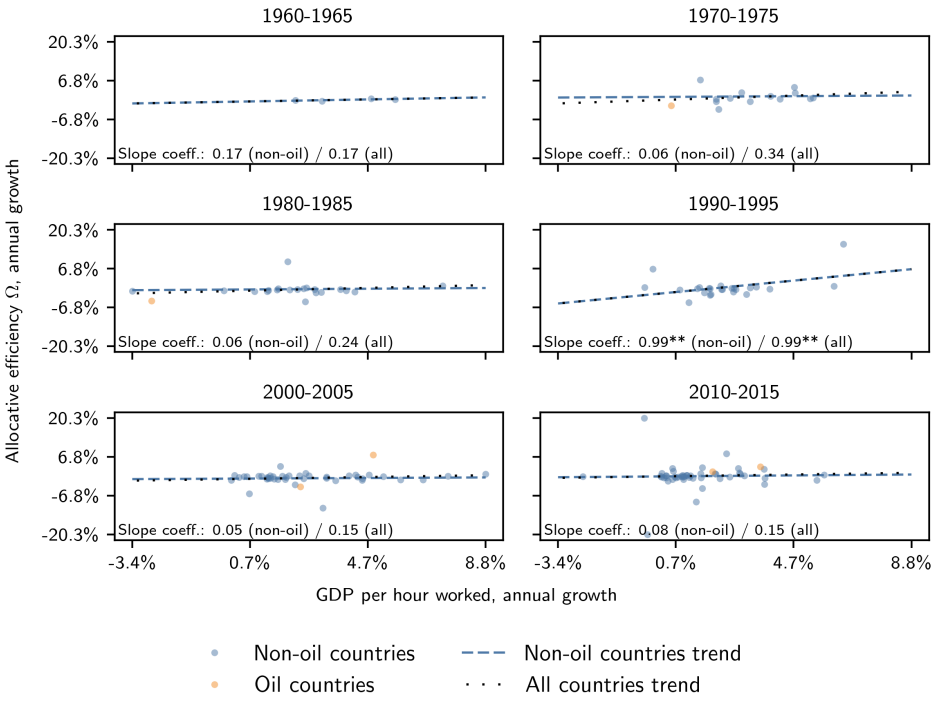


Figure G.12: Allocative efficiency growth vs. GDP per hour worked growth – growth dataset with best proxies ($k = 3$).

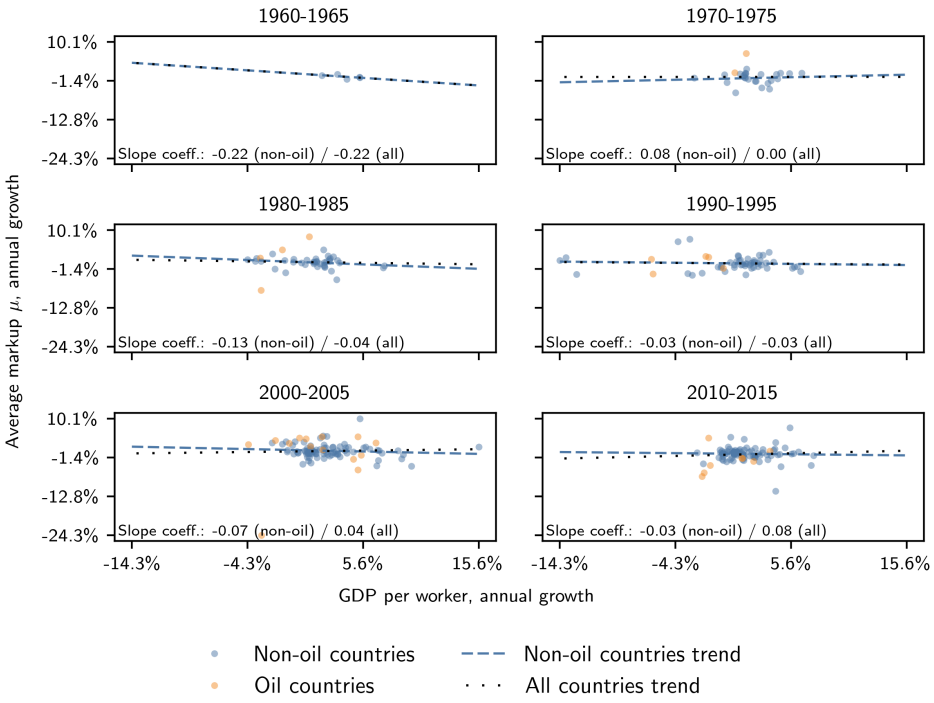


Figure G.13: Cost-weighted average markup growth vs. GDP per worker growth – growth dataset with broad coverage ($k = -1$).

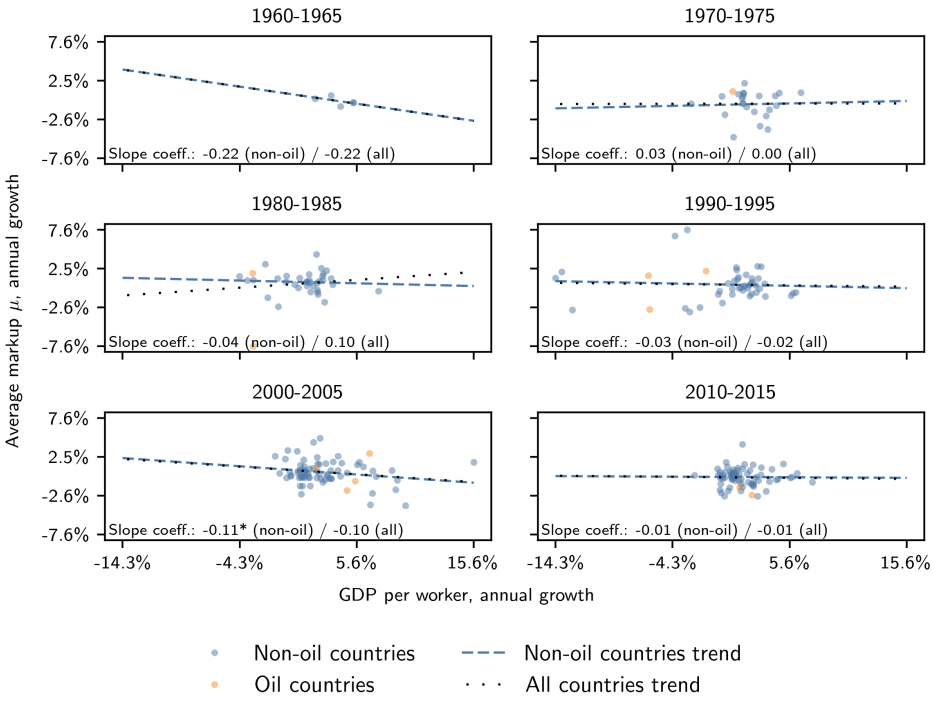


Figure G.14: Cost-weighted average markup growth vs. GDP per worker growth – growth dataset with broad coverage ($k = 3$).

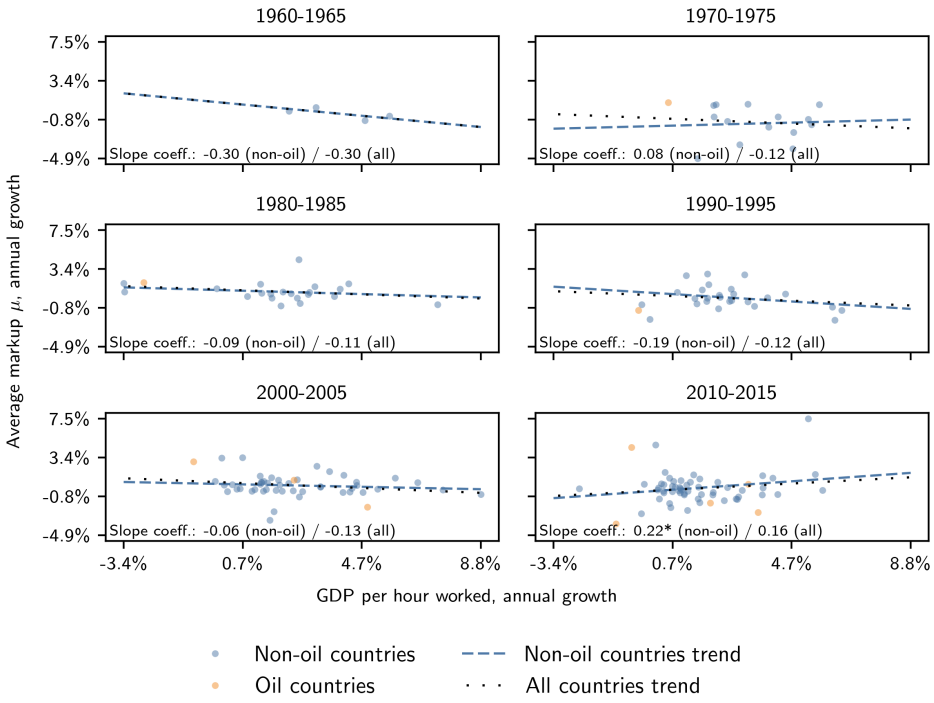


Figure G.15: Cost-weighted average markup growth vs. GDP per hour worked growth – growth dataset with best proxies ($k = -1$).

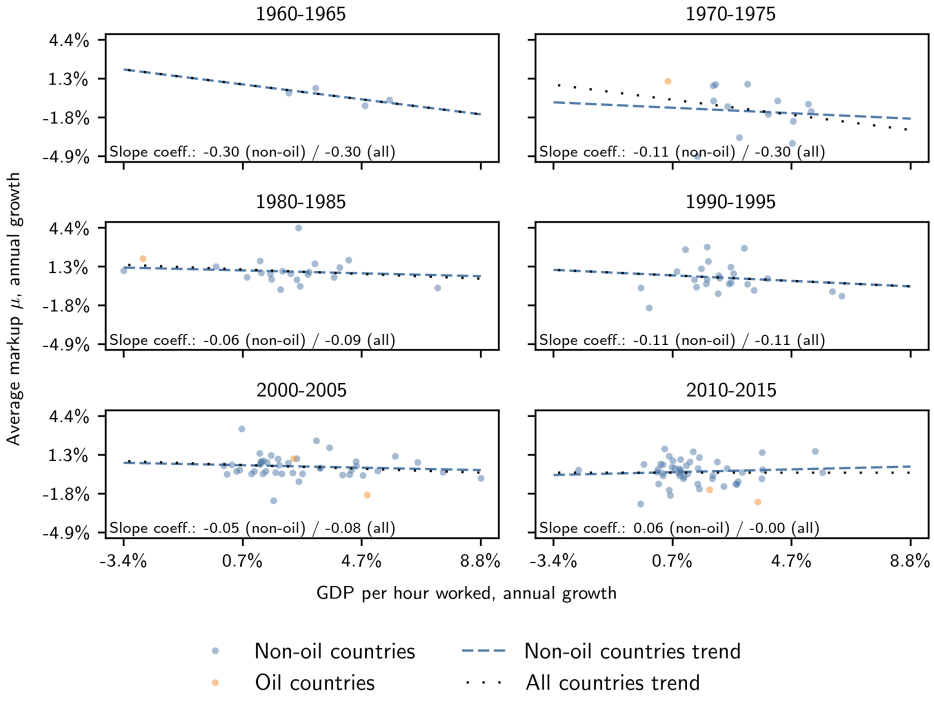


Figure G.16: Cost-weighted average markup growth vs. GDP per hour worked growth – growth dataset with best proxies ($k = 3$).

H Measures of success under common country samples

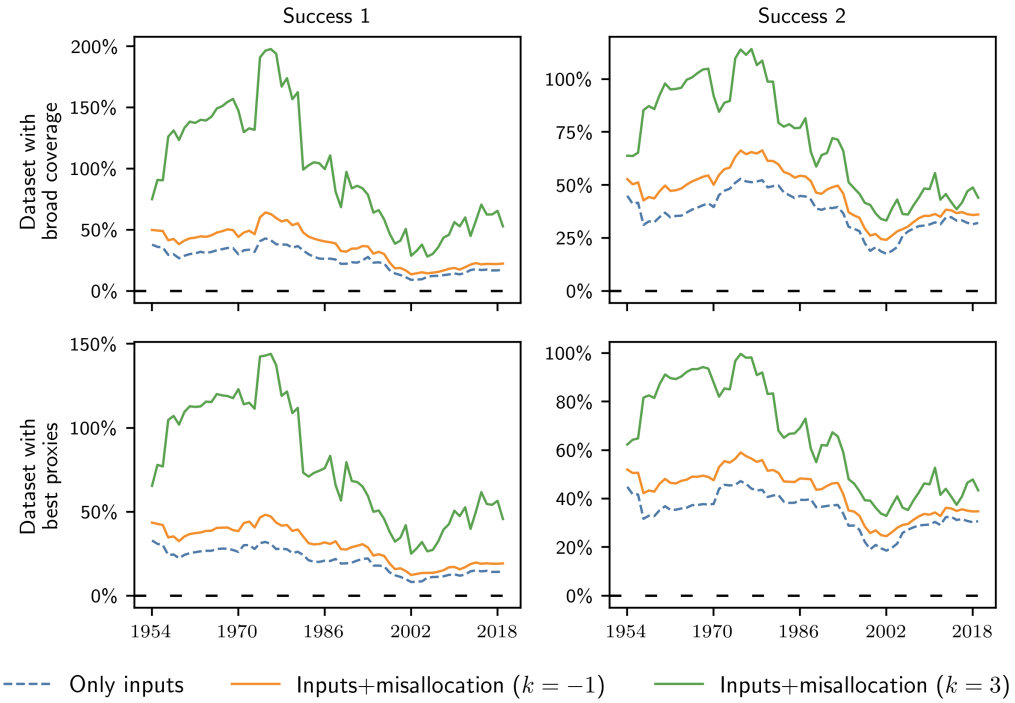


Figure H.1: Measures of success, common sample of non-oil countries across datasets for each year, 1954–2019.

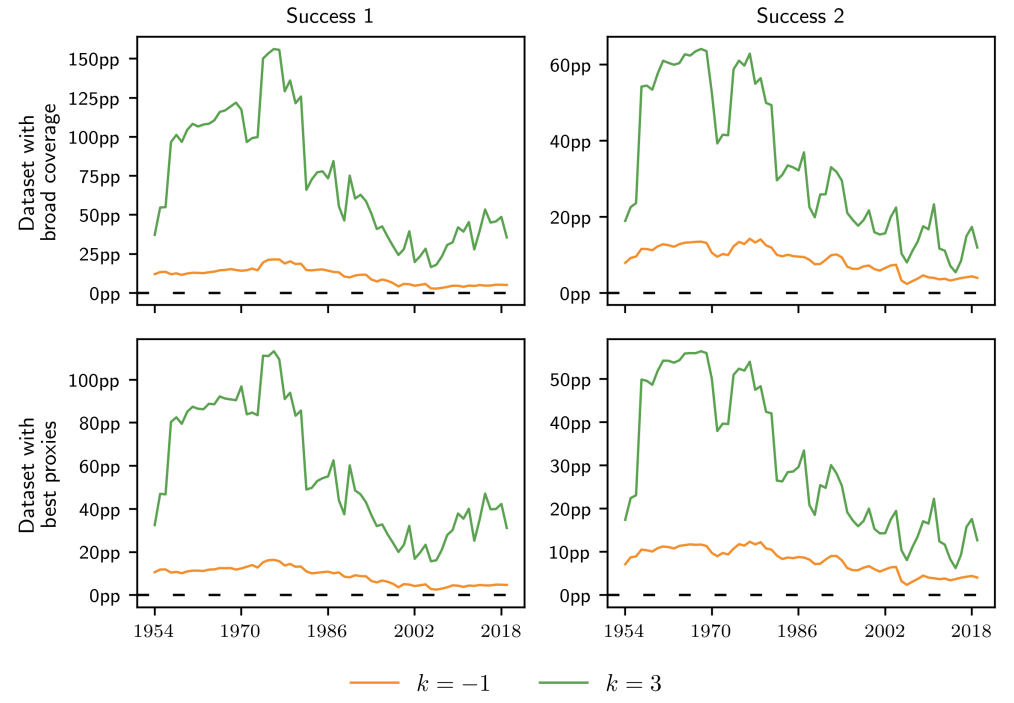


Figure H.2: Success gains due to misallocation, common sample of non-oil countries across datasets for each year, 1954–2019.

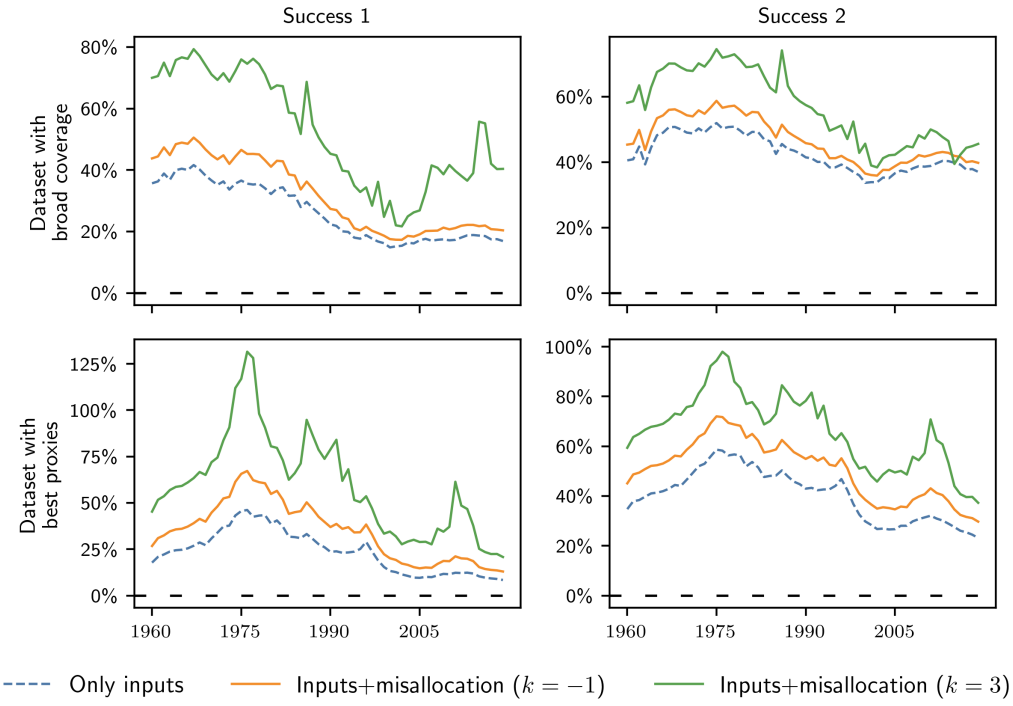


Figure H.3: Measures of success, common sample of non-oil countries over time for each dataset, 1960–2019.

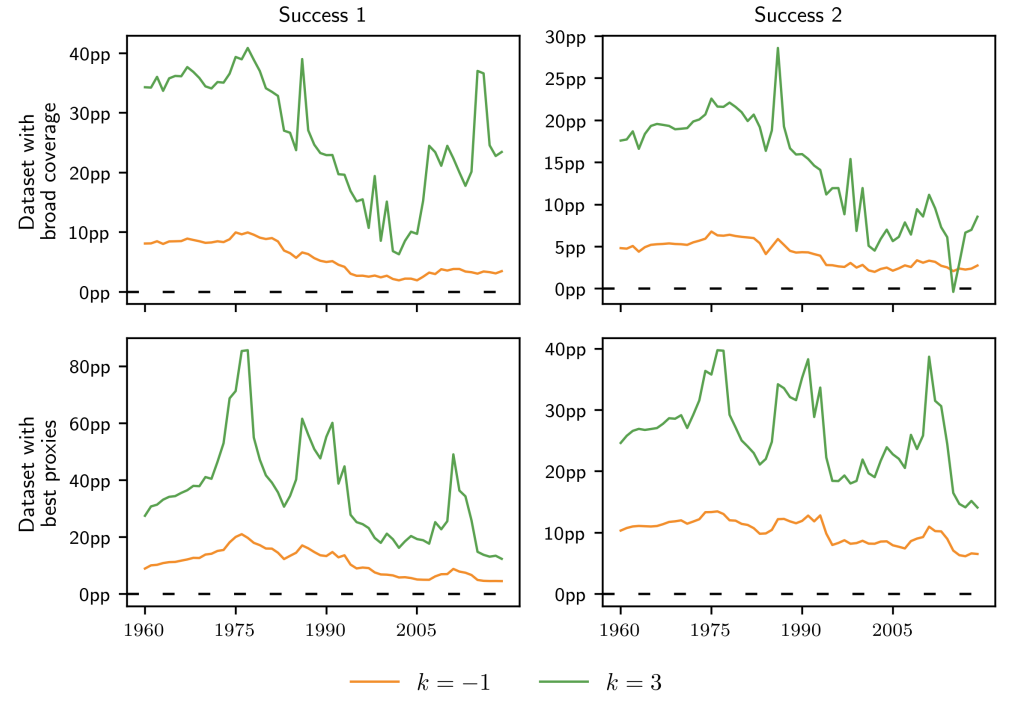


Figure H.4: Success gains due to misallocation, common sample of non-oil countries over time for each dataset, 1960–2019.

I Results under alternative calibrations of the capital elasticity

Table I.1: Growth accounting for the US ($\alpha = 1/3$)

Period	GDP/hour	GDP/hour decomp.			Labor-aug. TFP decomp.			
					$k=-1$		$k=3$	
	$\frac{Y}{E}$	$\left(\frac{K}{Y}\right)^{\frac{\alpha}{1-\alpha}}$	$\frac{L}{E}$	$\left(\overline{A}\Omega\right)^{\frac{1}{1-\alpha}}$	$\overline{A}^{\frac{1}{1-\alpha}}$	$\Omega^{\frac{1}{1-\alpha}}$	$\overline{A}^{\frac{1}{1-\alpha}}$	$\Omega^{\frac{1}{1-\alpha}}$
1954–2019	1.8	0.2	0.5	1.1	1.2	-0.1	1.2	-0.1
1954–1973	2.6	-0.0	1.0	1.7	1.7	0.0	1.7	0.0
1973–1995	1.3	0.3	0.5	0.5	0.7	-0.2	0.7	-0.2
1995–2000	2.5	0.2	0.3	2.0	1.3	0.7	1.1	0.9
2000–2007	2.1	0.5	0.3	1.3	1.9	-0.6	2.1	-0.7
2007–2019	1.0	0.2	0.2	0.6	0.7	-0.1	0.8	-0.2

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

Table I.2: Growth accounting for the US ($\alpha = 0.27$)

Period	GDP/hour	GDP/hour decomp.			Labor-aug. TFP decomp.			
					$k=-1$		$k=3$	
	$\frac{Y}{E}$	$\left(\frac{K}{Y}\right)^{\frac{\alpha}{1-\alpha}}$	$\frac{L}{E}$	$\left(\overline{A}\Omega\right)^{\frac{1}{1-\alpha}}$	$\overline{A}^{\frac{1}{1-\alpha}}$	$\Omega^{\frac{1}{1-\alpha}}$	$\overline{A}^{\frac{1}{1-\alpha}}$	$\Omega^{\frac{1}{1-\alpha}}$
1954–2019	1.8	0.1	0.5	1.1	1.2	-0.1	1.3	-0.1
1954–1973	2.6	-0.0	1.0	1.7	1.7	0.0	1.7	0.0
1973–1995	1.3	0.2	0.5	0.6	0.7	-0.2	0.9	-0.3
1995–2000	2.5	0.1	0.3	2.0	1.4	0.7	0.9	1.1
2000–2007	2.1	0.3	0.3	1.4	2.0	-0.5	2.4	-0.9
2007–2019	1.0	0.1	0.2	0.7	0.8	-0.1	0.9	-0.2

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

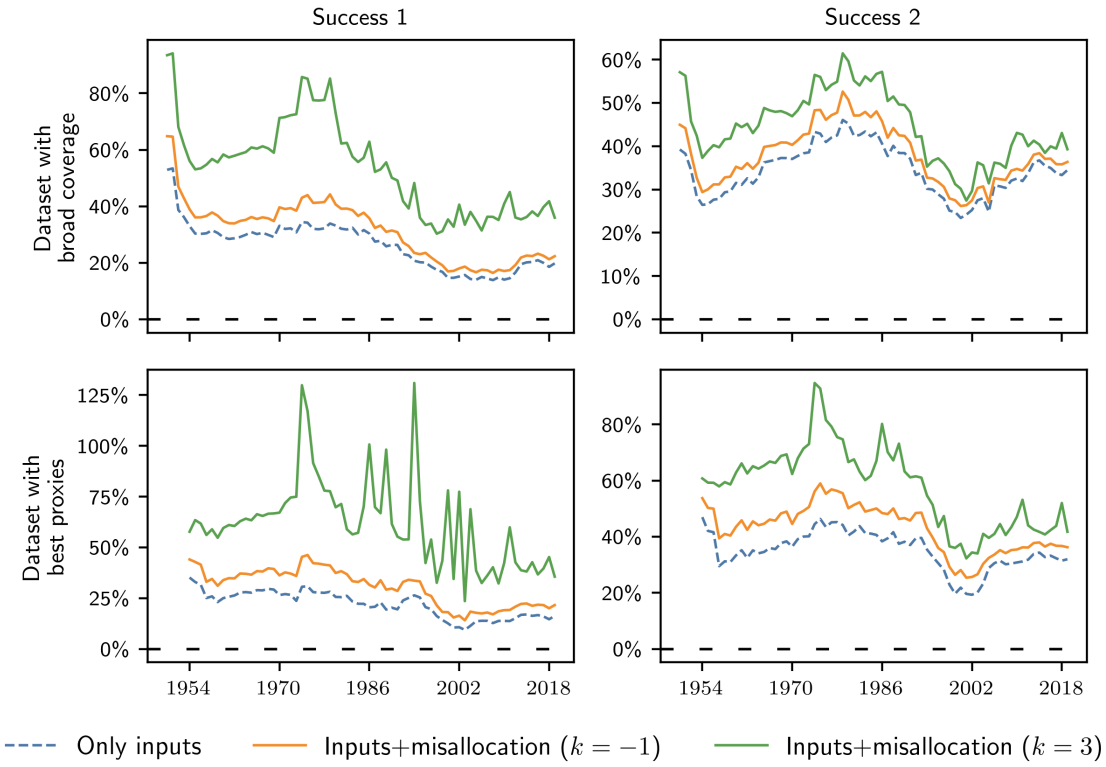


Figure I.1: Measures of success, common sample of non-oil countries for each dataset and year, 1950–2019 ($\alpha = 1/3$).

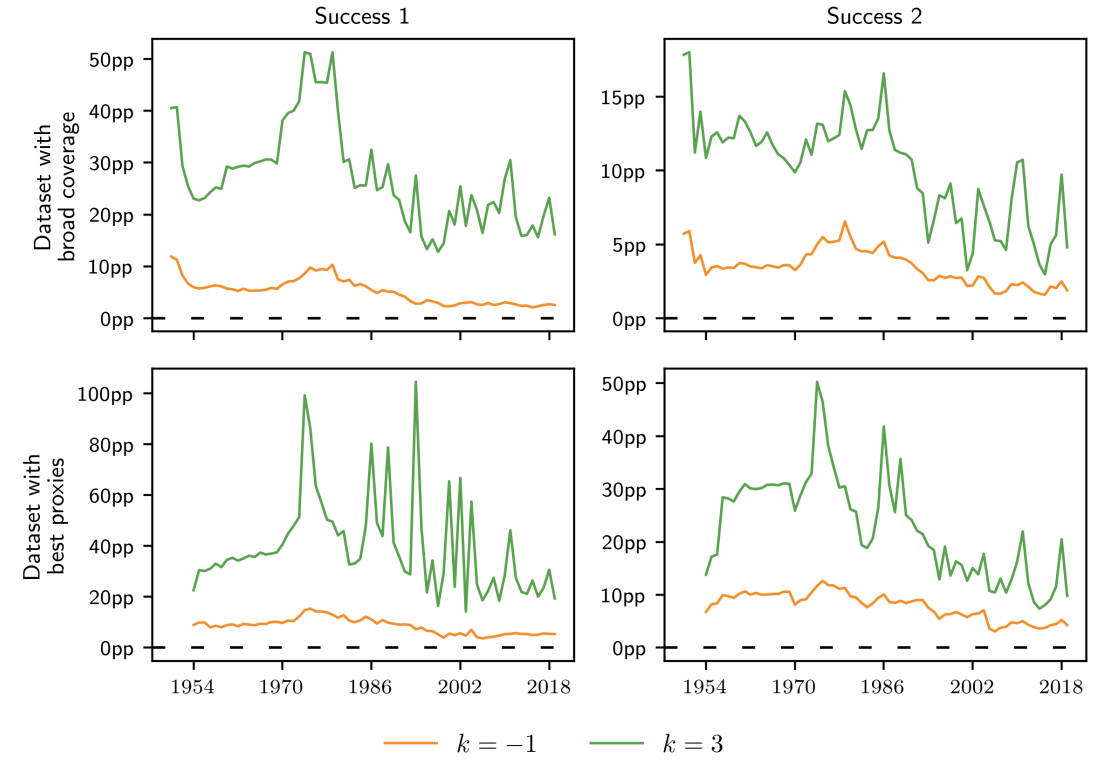


Figure I.2: Success gains due to misallocation, common sample of non-oil countries for each dataset and year, 1950–2019 ($\alpha = 1/3$).

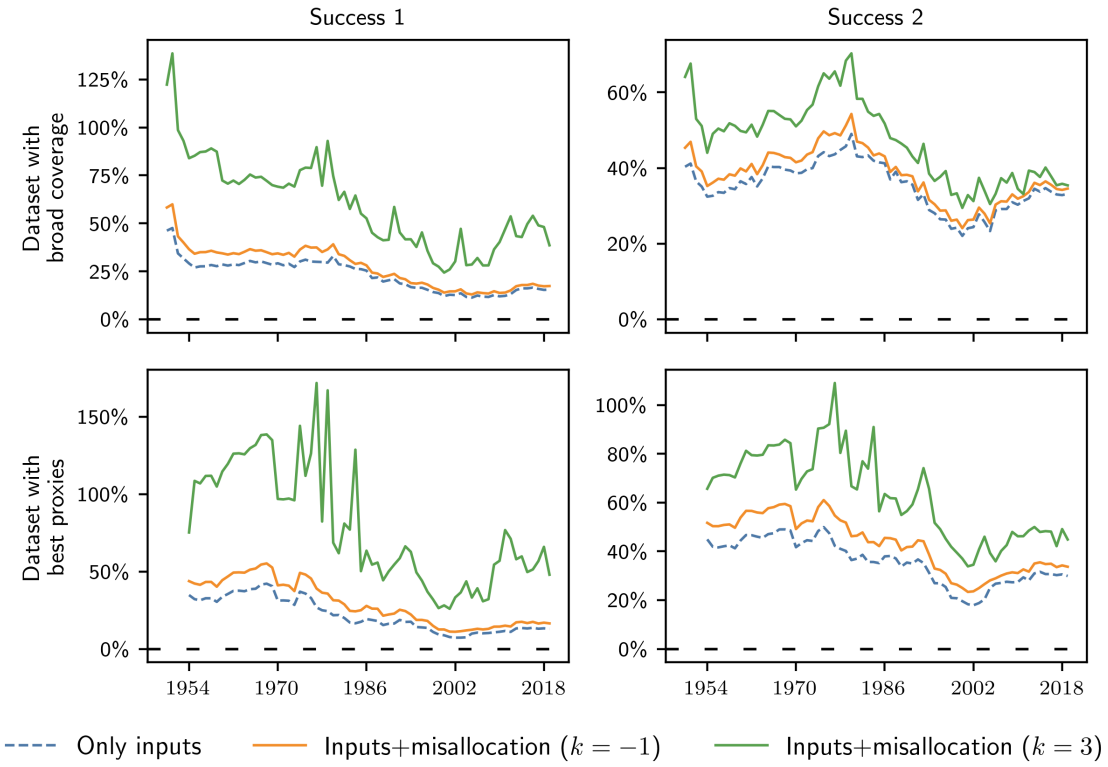


Figure I.3: Measures of success, common sample of non-oil countries for each dataset and year, 1950–2019 ($\alpha = 0.27$).

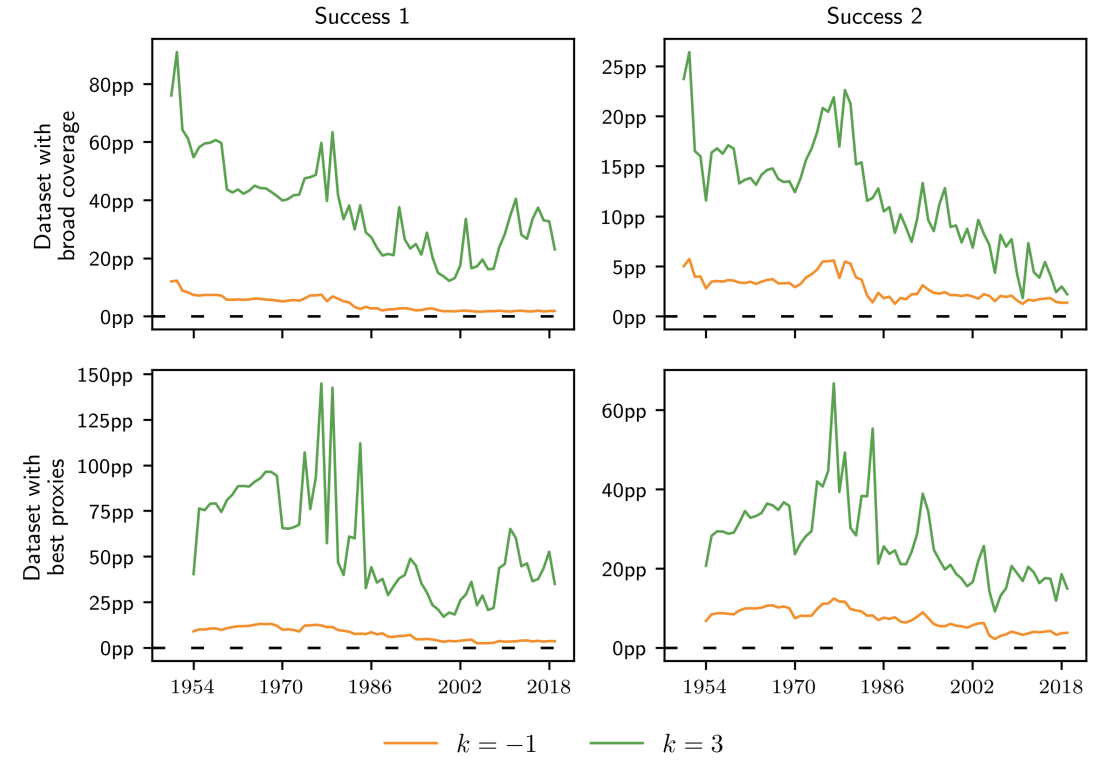


Figure I.4: Success gains due to misallocation, common sample of non-oil countries for each dataset and year, 1950–2019 ($\alpha = 0.27$).

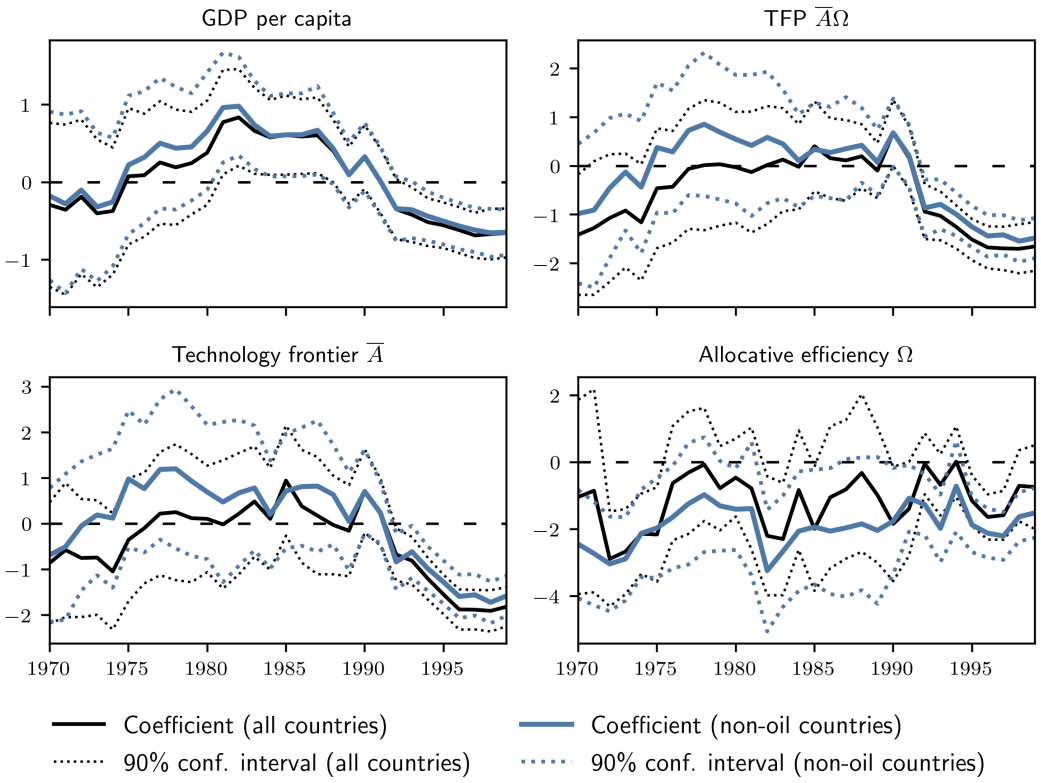


Figure I.5: Slope coefficient from regressing 20-year-ahead annual growth (in percent) on initial log levels, 1970–1999 – datasets with broad coverage ($k = -1$, $\alpha = 1/3$).

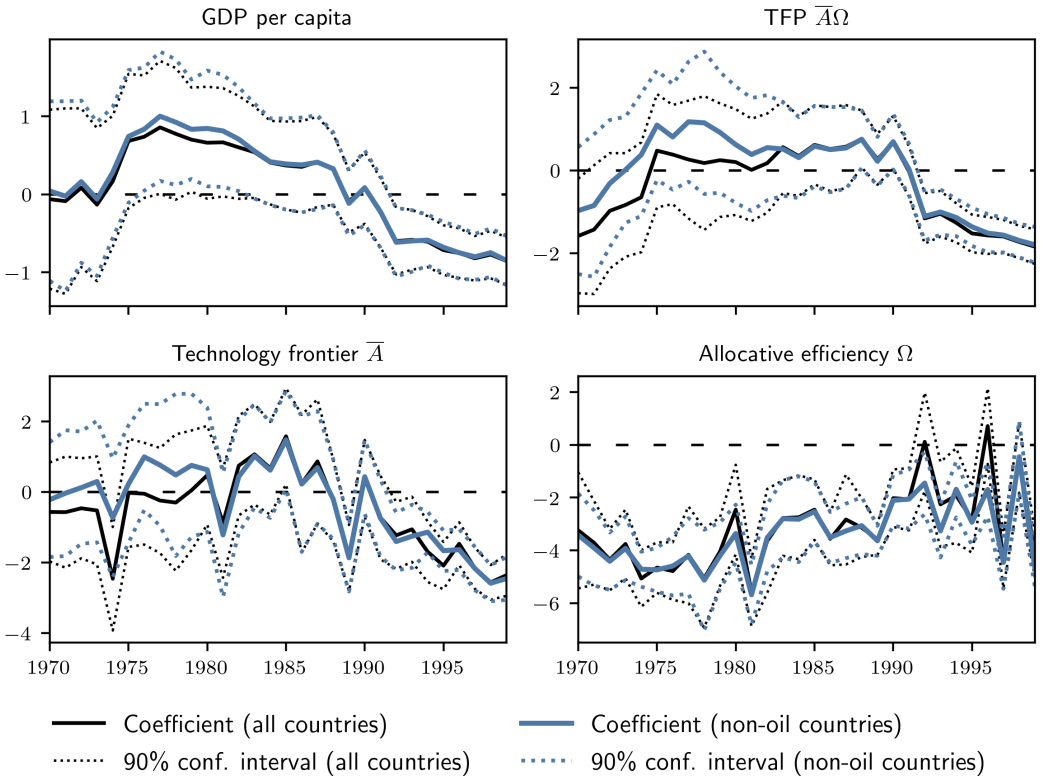


Figure I.6: Slope coefficient from regressing 20-year-ahead annual growth (in percent) on initial log levels, 1970–1999 – datasets with broad coverage ($k = 3$, $\alpha = 1/3$).

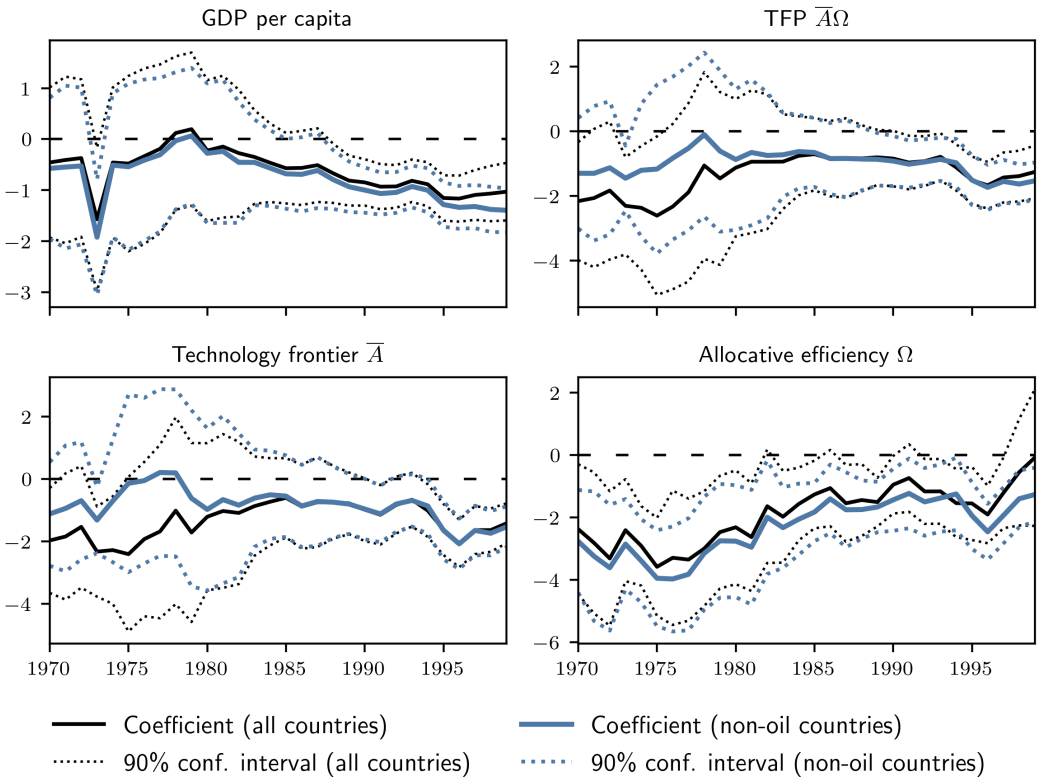


Figure I.7: Slope coefficient from regressing 20-year-ahead annual growth (in percent) on initial log levels, 1970–1999 – datasets with best proxies ($k = -1$, $\alpha = 1/3$).

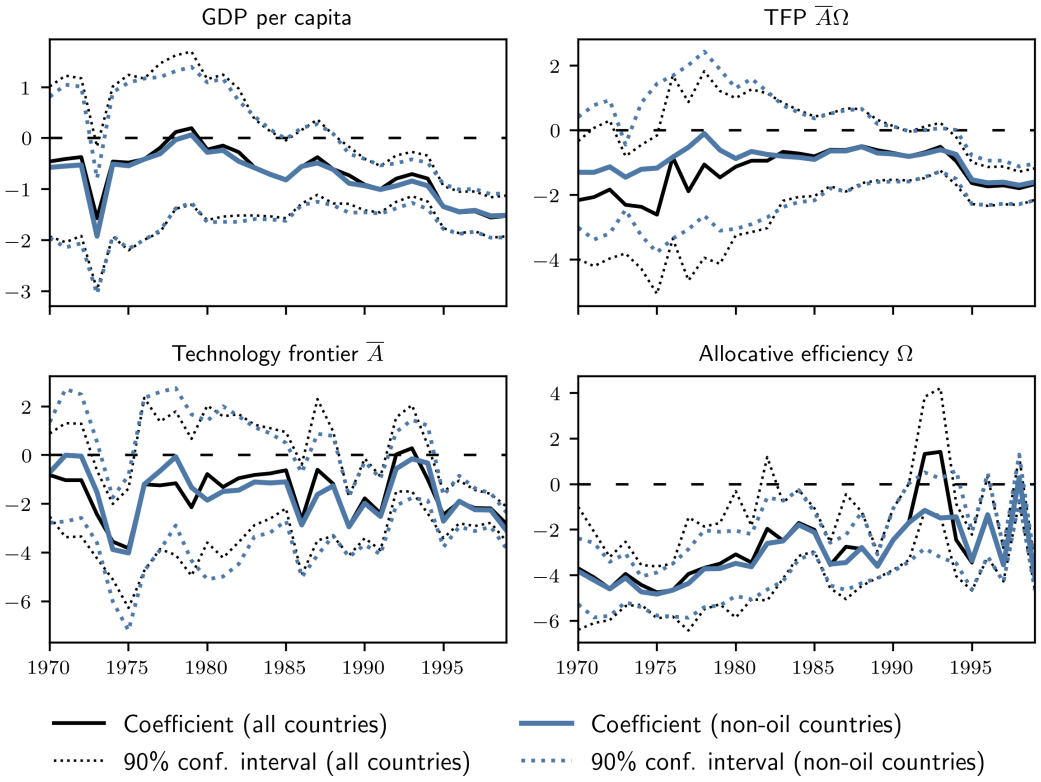


Figure I.8: Slope coefficient from regressing 20-year-ahead annual growth (in percent) on initial log levels, 1970–1999 – datasets with best proxies ($k = 3$, $\alpha = 1/3$).

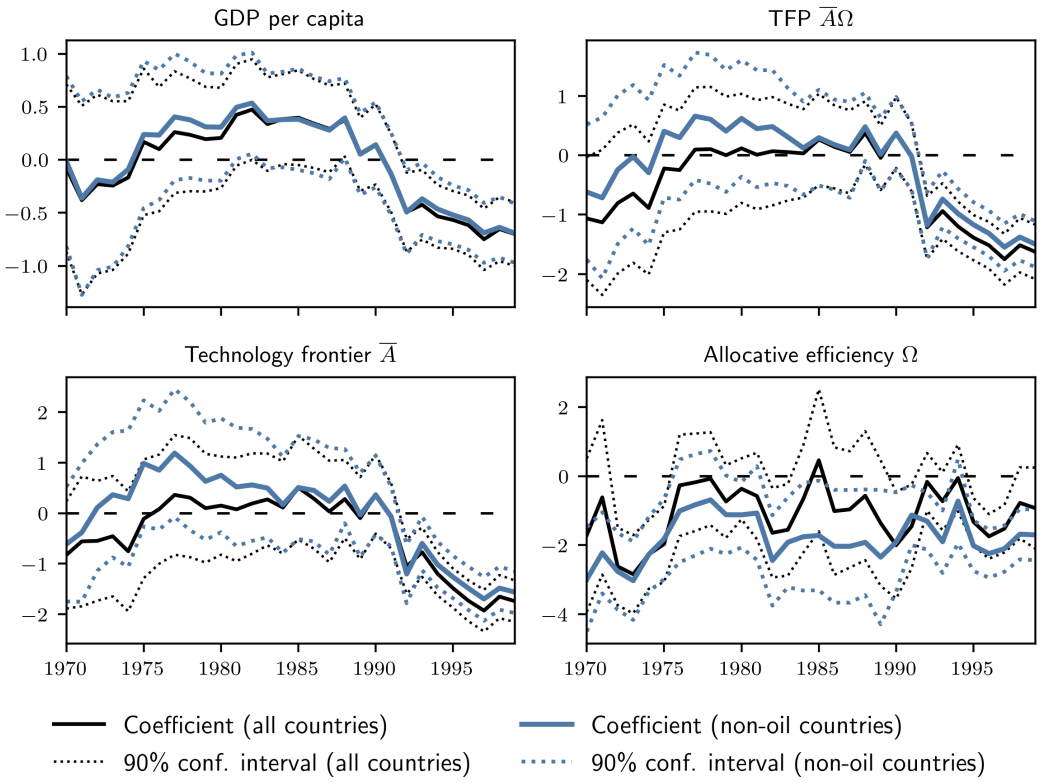


Figure I.9: Slope coefficient from regressing 20-year-ahead annual growth (in percent) on initial log levels, 1970–1999 – datasets with broad coverage ($k = -1$, $\alpha = 0.27$).

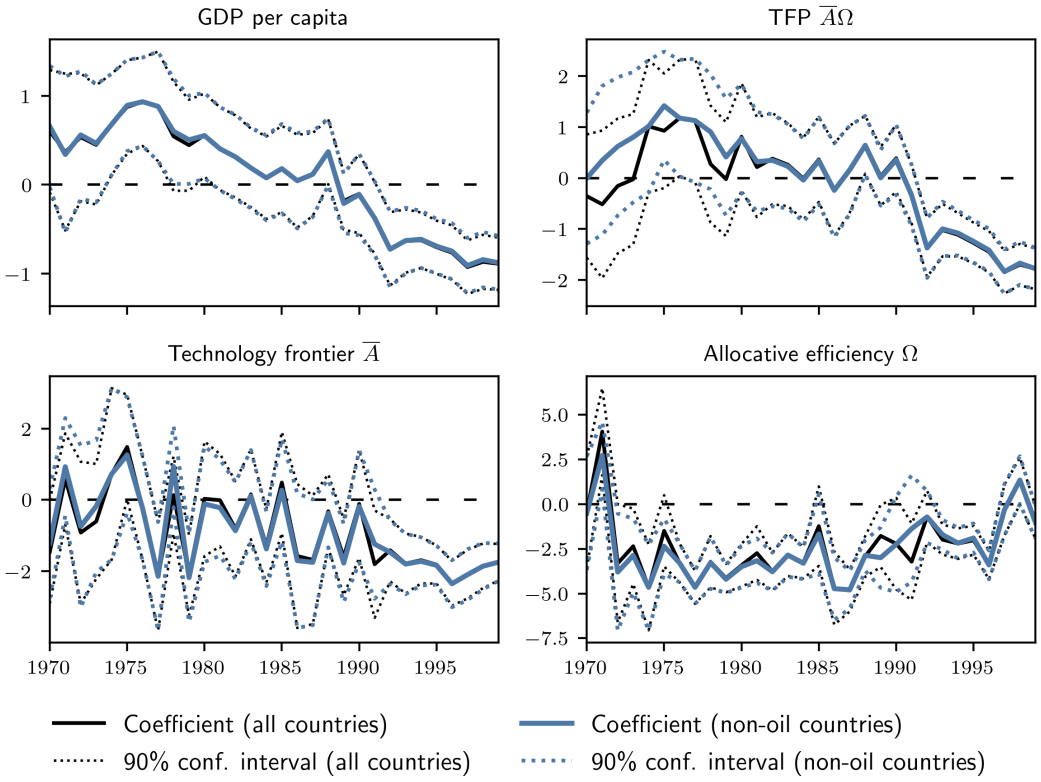


Figure I.10: Slope coefficient from regressing 20-year-ahead annual growth (in percent) on initial log levels, 1970–1999 – datasets with broad coverage ($k = 3$, $\alpha = 0.27$).

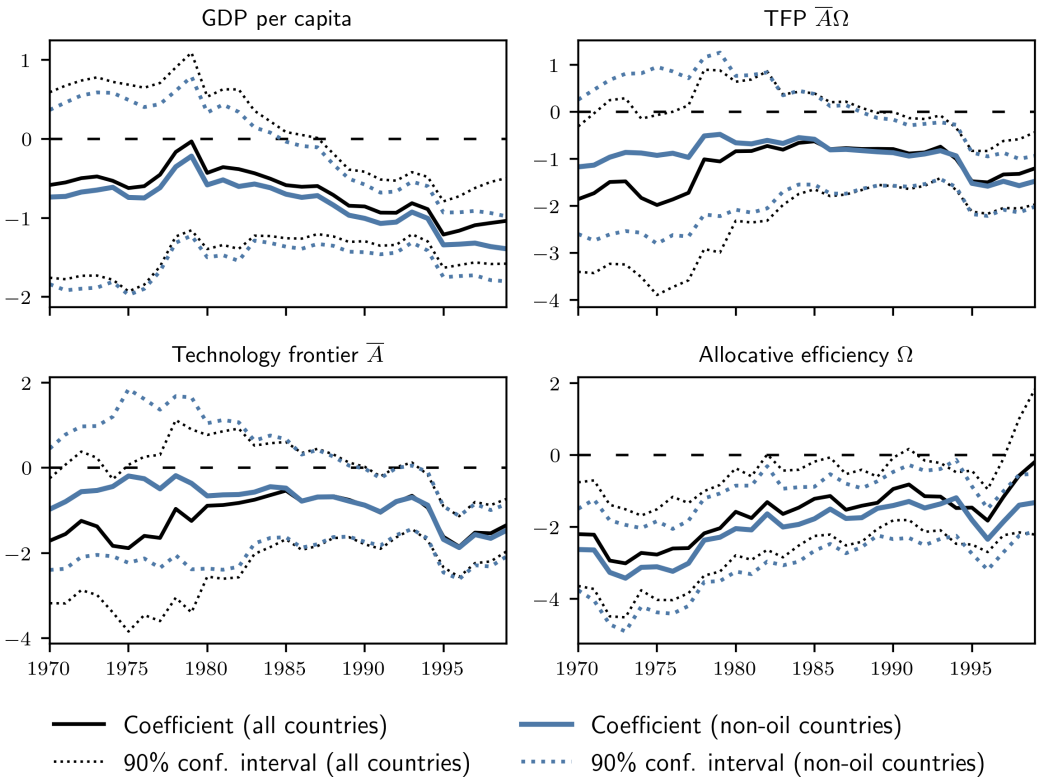


Figure I.11: Slope coefficient from regressing 20-year-ahead annual growth (in percent) on initial log levels, 1970–1999 – datasets with best proxies ($k = -1$, $\alpha = 0.27$).

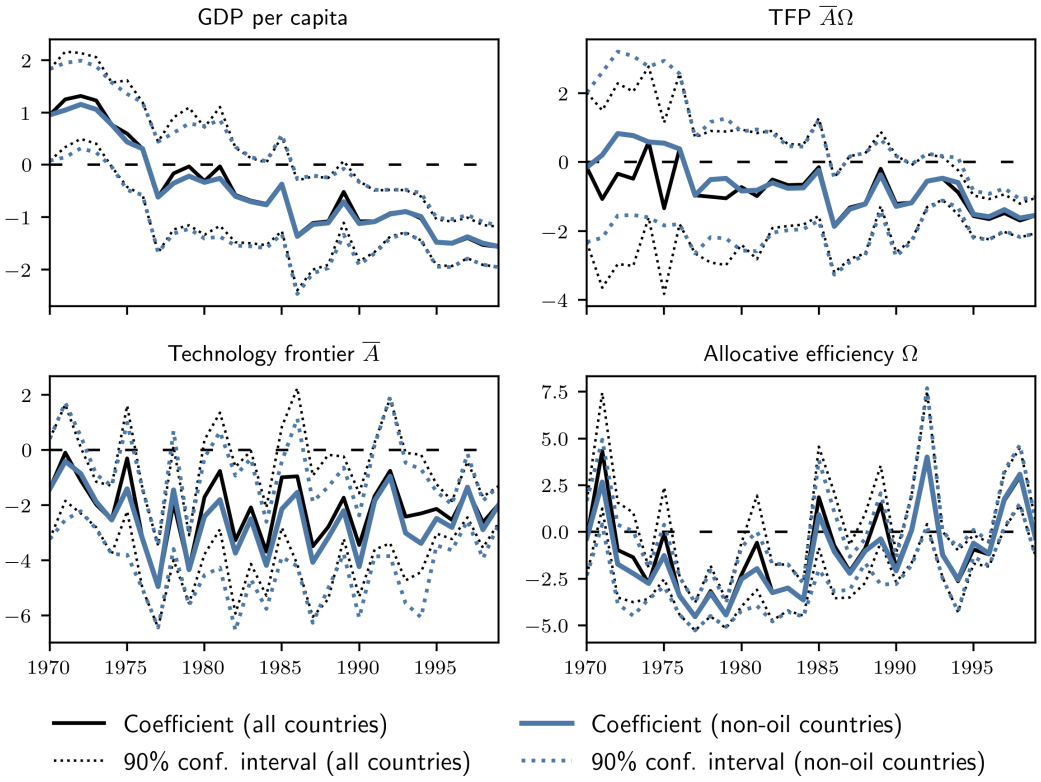


Figure I.12: Slope coefficient from regressing 20-year-ahead annual growth (in percent) on initial log levels, 1970–1999 – datasets with best proxies ($k = 3$, $\alpha = 0.27$).

J Measures of success for the full country sample

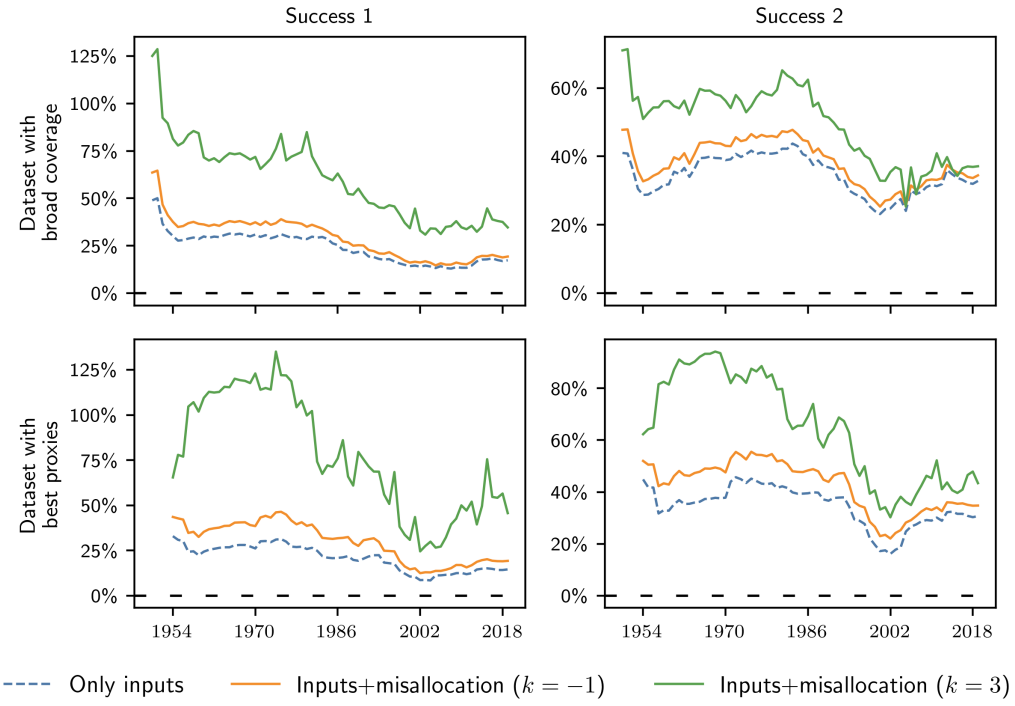


Figure J.1: Measures of success, common sample of countries for each dataset and year, 1950–2019.



Figure J.2: Success gains due to misallocation, common sample of countries for each dataset and year, 1950–2019.

References

Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2023). A theory of falling growth and rising rents. *Review of Economic Studies 90*(6), 2675–2702.

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. (2019). Are price-cost markups rising in the United States? a discussion of the evidence. *Journal of Economic Perspectives 33*(3), 3–22.

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.

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.

De Ridder, M. (2024). Market power and innovation in the intangible economy. *American Economic Review 114*(1), 199–251.

Díez, F. J., J. Fan, and C. Villegas-Sánchez (2021). Global declining competition? *Journal of International Economics 132*, 103492.

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 (2023). 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.

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 C. Santos (2026). Market power with formal and informal firms. *Mimeo*. Insper, and Research Department, Inter-American Development Bank, Washington, DC.

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

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.

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

Vollrath, D. (2024). The elasticity of aggregate output with respect to capital and labor. *American Economic Journal: Macroeconomics 16*(4), 470–504.

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