Aggregating Distortions in Networks with Multi-Product Firms

Yasutaka Koike-Mori — UCLA

Antonio Martner — UCLA & Central Bank of Chile*

[Preliminary version: June 9, 2024]

Abstract

We propose a theory of aggregating product-level distortions in a network economy and assess its implications for total factor productivity (TFP) growth. To this end, we provide a theoretical framework for growth accounting in inefficient economies with production networks and firms that engage in joint production. Using sufficient statistics summarizing inefficiencies arising from firm's product portfolio choices between products sold in different supply chains, we unpack between and within firm resource misallocation. This sufficient statistic is constructed as the covariance between the price change of a firm's product and a measure of the distortions accumulated in the downstream supply chain faced by that product. We apply the framework using a product transaction database (including granular prices and quantities) for the universe of formal Chilean firms. We find that within-firm allocative efficiency explains much of Chile's TFP growth after COVID-19 and in the subsequent high inflation period.

*We thank David Baqaee, Ariel Burstein, Hugo Hopenhayn, and Michael Rubens for excellent guidance. Pablo Fajgelbaum and Jonathan Vogel provided highly valuable comments. This study was developed within the scope of the research agenda conducted by the Central Bank of Chile (CBC) in the economic and financial affairs of its competence. The CBC has access to anonymized information from various public and private entities by virtue of collaboration agreements signed with these institutions. To secure the privacy of workers and firms, the CBC mandates that the development, extraction, and publication of the results should not allow the identification, directly or indirectly, of natural or legal persons. Officials of the Central Bank of Chile processed the disaggregated data. All the analysis was implemented by the authors and did not involve nor compromise the Servicios de Impuestos Internos de Chile. The views expressed are those of the authors and do not necessarily represent the views of the Central Bank of Chile or its board members.

1 Introduction

Prior work has shown the aggregate implications of firm-level distortions (Restuccia and Rogerson (2008), Hsieh and Klenow (2009)) with firm linkages (Baqaee and Farhi (2020)). These distortions go beyond technological progress and lead to changes in total factor productivity (TFP) due to shifts in resource allocation. This concept is known as allocative efficiency (AE).

While these papers acknowledge the importance of between-firm resource misallocation because they assume firms produce one product, they abstract from potential within-firm resource misallocation. If firms produce more than one product, the within-firm allocation, to which product they allocate their productivity resources, may also be distorted relative to the competitive allocation.

In reality, most firms engage in multi-product production; in Chile, we find that 72% of firms produce more than one product, often with different buyers for each product. This suggests that inefficiencies arising from the choice of product portfolios within firms could have a large impact on aggregate TFP.

Therefore, we build a theory to include within-firm AE and evaluate its contribution to aggregate TFP growth. We focus on markups (price over marginal cost) as the primary distortion, in line with existing literature. Relying on a granular database for Chile, including each product, price, and quantity transacted between formal firms in the economy, we find that within-firm AE accounts for around one-fourth of TFP growth in the last decade in Chile.

To theoretically describe within-firm AE, we extend Baqaee and Farhi (2020) production network structure by allowing firms to produce different products. Our theory assumes firms engage in joint production with a constant return to scale. This includes, as a special case, the separable production function by product, which is universally used in the literature. In the absence of joint production with multi-product firms, every product can be considered a separate firm, where shocks to one product do not affect the output of others. If every product is treated as a firm, then there is no within-firm misallocation, and we are back into the Baqaee and Farhi (2020) framework.

Specifically, to construct within-firm AE, we introduce the concept of "product network distortions," which captures firm-product distortions in terms of the entire production network of a country (rather than local distortions within a firm). These statistics summarize the relative magnitude of distortions accumulated through the network for each product. The allocation of a firm's within-product portfolio improves as it shifts toward more distorted products. We can show that this effect can be measured from sufficient statistics as covariance between product-specific price changes and the product-level network distortions. Furthermore, we can separate it from the between-firm AE and calculate the contribution of each to TFP.

To estimate within-firm AE, we need a granular product-level dataset. Leveraging on Chilean IRS sources, we use a decade of the universe of formal firms operating in Chile. Due to tax enforcement implications, every formal firm in Chile must declare all its invoices with other firms. The latter implies that for the universe of formal Chilean firms, we observe every product, quantity, and price traded with other firms. We also have access to tax accounting declarations from the Chilean IRS, which provides monthly data on each firm revenue and input expenditure, including capital and labor.

We bring our theory to the data to test the importance of within-firm AE. Our application involves building product-level production networks and recovering product-specific objects to compute aggregate TFP growth. We build product-level production networks using product flows between firms and estimate product-level network relative importance (centrality) measures. Building on Dhyne et al. (2022), we perform joint production function estimations, computing product-specific marginal costs and markups.

We compute product-level network distortions from product-level sales and markups and use them to compute within-firm and between-firm AE. We then decompose the TFP growth as technological changes, between-firm AE and within-firm AE. We find that in Chile, while the main component explaining TFP growth is the allocation of resources between firms, within-firm AE (how firms choose their product portfolio) accounts for around one-fourth of TFP growth for the last decade in Chile.

Although our theory on within-firm AE applies to any economy, our application to Chile provides a potential explanation for a longstanding puzzle: the stagnation of TFP growth in Chile since the early 2010s. Our findings attribute the observed TFP growth to changes in AE, both between and within firms, rather than technological change. While our insights are specific to Chile, they also hold broader implications that may apply to other countries. Chile's macroeconomic experience mirrors global productivity dynamics, characterized by a surge in TFP before the Great Recession, followed by a period of stagnation.

Related Literature

This paper is related to the literature on misallocation in production networks and growth accounting. The misallocation literature has been developing both empirically and theoretically in since seminal works by Restuccia and Rogerson (2008); Hsieh and Klenow (2009), and recently there has been an extensive study of inefficiencies in the allocation of factors of productivity in the presence of firm linkages (Bigio and La'O (2020); Osotimehin and Popov (2023); Dávila and Schaab (2023); Liu (2019); Baqaee and Farhi (2020); Baqaee et al. (2023)).

We contribute to the literature both theoretically and empirically. On the theoretical side, we allow for joint production structures within production networks and develop a general framework to detect the contribution of allocative efficiency to TFP. Our theory provides a tool for growth accounting (Solow (1957); Hulten (1978); Basu and Fernald (2002); Petrin and Levinsohn (2012)) that decomposes TFP calculated from observed data into three interpretable terms, which are technology, between-firm AE, and between-firm AE. If firms engage in joint production, existing approaches cannot be applied by treating different products as different firms. Moreover, even if we apply existing methods under the assumption of non-joint production, it is not possible to separate between and within-firm AE, as we do. In this sense, we generalize existing methods and provide a way to maximize data mapping to a theoretical model.

Our second contribution is to apply our theory using a product-level trade database covering a single country to quantify misallocation. Much of the prior literature on networks and misallocation uses industry-level input-output tables instead of firm-to-firm networks. For example, Baqaee and Farhi (2020) imputes U.S. Compustat data with an industry-level input-output table. Liu (2019) examines the role of wedges in industry-level IO linkages. Even when firm-level transaction data are available, the lack of complete price information limits the analysis. For example, Kikkawa (2022) examines firm pair-specific markups based on a theoretical model using Belgian interfirm transaction data. In contrast, we developed and implemented a methodology to measure misallocation using a dataset that covers the universe of firm-to-firm transactions at the product level.

Our study is also tangentially relevant to the literature on production function estimations. Particularly, estimation methods for joint production have been developed recently Dhyne et al. (2017, 2022); De Loecker et al. (2016); Valmari (2023); Cairncross and Morrow (2023). Even though we do not develop any theoretical innovation, unlike previous papers that estimate production functions for a specific industry or subsample of the economy, we apply Dhyne et al. (2022) method to estimate multi-product production functions for the universe of products traded in Chile by formal firms for the 2014-2022 period.

2 A theory to aggregate product-level inefficiencies

We propose a flexible theoretical framework for maximizing data mapping in an inefficient multi-product firm input-output structure. Specifically, we characterize the first-order effects of firm-product-level shocks on aggregate TFP in an economy with arbitrary wedges and product-level production networks. This framework generalizes Hulten (1978) and Baqaee and Farhi (2020) single-product firm production network theory to a multi-product production network environment.

Although our setup is valid for any wedge, we will assume markups are the sole one because we can rely on Dhyne et al. (2022) theory to estimate product level markups derived from multi-product joint production firms. The latter enables us to assess how markups generate production factors misallocation between and within firms, shaping TFP growth.

2.1 Joint Production

In a joint production setup, firms use common inputs to produce a product portfolio, meaning that some inputs may simultaneously be used in the production of multiple products. Joint production implies that shocks to one product might affect the output of others as they share common production inputs. In the absence of joint production with multi-product firms, every product can be considered a firm. If every product is treated as a firm, then there is no within-firm misallocation, and we are back into Baqaee and Farhi (2020) framework.

Before moving to the network setup, we will discuss joint production in general. In a joint production setup following Diewert (1973), let $t(\boldsymbol{q}, \boldsymbol{x})$ as transformation function:

$$V(\boldsymbol{q}) = \{\boldsymbol{x} | t(\boldsymbol{q}, \boldsymbol{x}) \geq 0\},$$

where $\boldsymbol{q}$ is the vector of outputs and $\boldsymbol{x}$ is the vector of inputs. The cost function is based on cost minimization, as follows:

$$C(\boldsymbol{q}, \boldsymbol{p}) \equiv \min_{x \in V(q)} \boldsymbol{p} \times \boldsymbol{x}, \tag{1}$$

We introduce assumptions about the shape of some production sets. These two assumptions will be used throughout this paper.

Assumption 1. constant return to scale: $t(\boldsymbol{q}, \boldsymbol{x}) = 0$ imples $t(\lambda \boldsymbol{q}, \lambda \boldsymbol{x}) = 0$.

The difference from a production function with a single output is that the output is a vector.

Assumption 2. Separability between input and output function, Hall (1973): $t(\boldsymbol{q}, \boldsymbol{x}) = -g(\boldsymbol{q}) + f(\boldsymbol{x})$ has the joint cost function: $C(\boldsymbol{q}, \boldsymbol{p}) = H(\boldsymbol{q}) \varphi(\boldsymbol{p})$.

Note that this is different from assuming non-joint production functions when a firm has more than two products. In that case, the output $\boldsymbol{q}$ of in $g(\boldsymbol{q})$ is not a vector but a single product and thus degenerates to $g(q) = q$.

We illustrate a production set of classes satisfying Assumptions 1 and 2:

Example 1. Constant-Elasticity of Transformation and CES Input (CET-CES):

$$\underbrace{\left(\sum_{g} q_{ig}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}}_{\text{Output bundle}} = A \underbrace{\left(\omega_L L^{\frac{\sigma-1}{\sigma}} + \omega_K K^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}}_{\text{Input Bundle}}$$

where L and K are the two inputs and q is a vector of outputs.

The input bundle take a standard CES function, but the output bundle is a vector of goods rather than a scalar. This functional form is discussed in Hall (1988), and the parameter σ is called the elasticity-constant transformation elasticity, which gives a constant value to the PPF curvature of the products within a firm. However, since our purpose is an ex-post exercise, our results below do not require such parametric assumptions.

2.2 Network Setup

We use Baqaee and Farhi (2020) input-output notation and definitions to present our extension and add product-level (instead of firm-level) objects. We map how product-level markups spread throughout production networks, which determine distorted within-firm factor allocations.

Multi-Product Firms

Each firm $i \in I$ can produce different products $g \in G$ combining inputs (products) from other firms $j \in J$ with production factors[1] (Labor, L and Capital, K) following the transformation function.

Firm $i \in N$ produces product $g \in G$ and use products $p \in G$ from another firm j and factors as inputs. We assume production set with CRS and separability

$$F_i^Q\left(\left\{\underbrace{q_{ig}}_{\text{outputs}}\right\}_{i\in N, g\in G}\right) = A_i F_i^X\left(\left\{\underbrace{x_{i,jp}}_{\text{Intermediate product } p \text{ from j}}\right\}_{j\in N, p\in G}, L_i, K_i\right),$$

Firms charge prices, including a product-specific markup μ_{ig} over its marginal cost $p_{ig} = mc_{ig} \times \mu_{ig}$.

Final Demand

There is a representative household with utility function constant-return to scale utility function $U\left(c_{ig}, ..., c_{NG}\right)$ that receives firm profits through an income transfer T and faces the following budget constraint:

$$\sum_{i\in N}\sum_{g\in G} p_{ig}c_{ig} = \sum_{f\in\{L,K\}} w_f L_f + \sum_{i\in N}\sum_{g\in G}\left(1 - 1/\mu_{ig}\right) p_{ig} q_{ig} + T$$

Each product can either be consumed by the final consumers (c_{ig}) or used as an input in production by other firms ($x_{ji,g}$), facing the following resource constraints:

$$q_{ig} = c_{ig} + \sum_{j\in N} x_{jig} \quad \sum_{i\in N} L_i = L \quad \sum_{i\in N} K_i = K$$

2.2.1 General Equilibrium

Given productivity A_i and markup μ_{ig} for all $i \in N$ and $g \in G$, the general equilibrium is the set of prices p_{ig}, intermediate input choices $q_{ijg'}$, factor input choices $\{L_i, K_i\}$, output q_{ig}, and consumption choices c_{ig}, such that: (i) the price of each good is equal to its

[1]We assume factors exhibit zero return to scale production functions, in a sense, they generate inputs without using inputs from other firms.

markup multiplied by its marginal cost; (ii) households maximize utility under budget constraints, given prices; and (iii) markets clear for all goods and factors.

National Account

GDP is defined as the sum of all product values consumed by the final consumers: $GDP = \sum_{i\in N}\sum_{g\in G} p_{ig}c_{ig}$

Real GDP (Y) changes can be computed as

$$d\log Y = d\log GDP - \sum_{i\in N}\sum_{g\in G}\frac{p_{ig}c_{ig}}{GDP}d\log p_{ig}$$

Factor Shares are defined as

$$\Lambda_L = \frac{wL}{GDP}, \quad \Lambda_K = \frac{rK}{GDP}$$

Product-Input-Output Networks

The product-input-output matrix $\tilde{\Omega}$ is a $(N \times G + F)$ square matrix where N is the number of firms, G is the number of products, and F is the number of factors. $\tilde{\Omega}$ has at its ig, jp^{th} element the expenditure share of product p from firm j and factor $f \in F$ used by firm i in production over firm i total costs (of producing all its goods). From the separability assumption, the same expenditure share applies to all products, g that i makes. Thus, $\tilde{\Omega}_{ig,jp}$ and $\tilde{\Omega}_{ig,f}$ are as follows.

$$\tilde{\Omega}_{ig,jp} = \frac{p_{jp}x_{i,jp}}{\sum_{j,p} p_{jp}x_{i,jp} + \sum_f w_f L_{if}}, \quad \tilde{\Omega}_{ig,f}=\frac{w_f L_{if}}{\sum_{j,p} p_{jp}x_{i,jp} + \sum_f w_f L_{if}}$$

The product cost-based Leontief inverse $\tilde{\Psi}$ captures every firm-product pair's direct and indirect cost exposures through production networks. Each element of $\tilde{\Psi}$ measures the weighted sums of all paths (steps) between any two non-zero firm-product pairs.

$$\tilde{\Psi} \equiv (I - \tilde{\Omega})^{-1} = I + \tilde{\Omega} + \tilde{\Omega}^2 + \dots$$

To measure the relative importance of each firm-product pair on GDP, we define GDP shares in vector form:

$$b_{ig} = \begin{cases} \frac{p_{ig}c_{ig}}{GDP} & \text{if } i \in N, g \in G \\ 0 & \text{otherwise} \end{cases}$$

We set GDP to be the numeraire, and we define the product level cost-based Domar weight vector[2] $\tilde{\lambda}_{ig}$ measures the importance of product g from firm i in final demand in two dimensions, directly when it is sold to final consumers and indirectly through the production network when it is sold to other firms that eventually, downstream production networks will reach final consumers.

$$\tilde{\lambda}' \equiv b'\tilde{\Psi} = b' + b'\tilde{\Omega} + b'\tilde{\Omega}^2 + \dots$$

Firm Level Aggregation

Summing over products by firms, we recover the firm-level cost-based Domar weight $\tilde{\lambda}_i$, which we use to compute the within-firm product-level Domar weight share s_{ig}:

$$\tilde{\lambda}_i = \sum_{g \in G} \tilde{\lambda}_{ig}, \quad s_{ig} = \frac{\tilde{\lambda}_{ig}}{\tilde{\lambda}_i} \tag{2}$$

Finally, we define firm-level aggregate markup as

$$\mu_i = \frac{\text{sales of firm } i}{\text{total cost of firm } i} = \left[\frac{\sum_{g \in G} mc_{ig} y_{ig}}{\sum_{g \in G} \frac{1}{\mu_{ig}} mc_{ig} y_{ig}} \right]^{-1},$$

Product Distortion in networks

With all the needed ingredients, we define the main object of our theory: "product network distortion" Γ_{ig}, which measures each firm-product pair distortion through production networks. It is cost-based Domar weight, $\tilde{\lambda}_{ig}$ to the product-level ratio of the total cost, $sales_{ig}/\mu_{ig}$. In an efficient economy, $sales_{ig}/\mu_{ig} = \tilde{\lambda}_{ig}$, but in inefficient economies, they are different.

[2] We use $\tilde{\Lambda}_f$ with $f \in \{L, K\}$ $\tilde{\lambda}'$ of length $(N \times G + 2) \times 1$.

Definition 1. Product Network Distortion

$$\Gamma_{ig} \equiv \frac{\tilde{\lambda}_{ig}}{sales_{ig}/\mu_{ig}} = \frac{\text{Cost-based Domar weight of product } g \text{ made by firm } i}{\text{Total cost of } g \text{ made by firm } i}$$

Intuitively, this statistic measures the size of the inefficiency of the buyer of the product and its downstream supply chain. The more distorted the supply chain, the larger the size of the sales will be relative to the optimal allocation, so the larger this value means that the supply chain that this product is exposed to is underproducing.

Next, we define the relative product network distortion, which ranks product distortions within a given firm. It is measured as the relative downstream distortion of product g with respect to the average distortion of firm off all products within firm i, Γ_i.

Definition 2. Relative Product Network Distortion

$$\gamma_{ig} \equiv \frac{\Gamma_{ig}}{\Gamma_i}$$

where average distortion of firm i is $\Gamma_i \equiv \sum_g \tilde{\lambda}_{ig} / \sum_g \left(sales_{ig}/\mu_{ig}\right)$.

γ_{ig} describes relative distortion differences within a firm and will be the object that will determine within-firm allocative efficiency. If $\gamma_{ig} > 1$, it means that product g is more distorted relative to firm i mean distortion.

Value of γ_{ig}	relative distortion within firm i
> 1	more distorted
$= 1$	average distortion
< 1	less distorted

Table 1: Interpretation of relative product network distortion, γ_{ig}

2.3 Growth accounting

In this section, we generalize the concept of an allocation matrix introduced by Baqaee (2020) to allow for joint production and derive sufficient statistics. This is useful in growth accounting because it provides an interpretable decomposition from the observables.

Let $\mathcal{X}^{IN}$ be an $(N+F)\times(N\times G+F)$ admissible input allocation matrix, where the columns are buyer firms and the rows are seller-product pairs. Each of its elements $\mathcal{X}^{IN}_{ijg} = \frac{q_{ijg}}{q_{jg}}$ is the share of the output of good g produced by firm j that firm i uses as an input.

For each firm i, we define the output correspondence $\mathcal{X}^{OUT}_i$. This output correspondence $\mathcal{X}^{OUT}_i$ maps the rows of the matrix $\mathcal{X}^{IN}_i$, corresponding to firm i, to a vector of outputs for each good produced by firm i. This vector can be represented as $\boldsymbol{q}_i = (q_{i1}, q_{i2}, \ldots, q_{iG})$, where q_{ig} is the output of good g produced by firm i. Formally, the output correspondence can be defined as:

$$\mathcal{X}^{OUT}_i : \mathcal{X}^{IN}_i \mapsto \boldsymbol{q}_i$$

The set of all input allocation matrices and their corresponding output correspondences for each firm is defined as a collection $\mathcal{X}$, which can be formally represented as:

$$\mathcal{X} = \left(\mathcal{X}^{IN}, \mathcal{X}^{OUT}\right)$$

where $\mathcal{X}^{OUT} = \{\mathcal{X}^{OUT}_1, \mathcal{X}^{OUT}_2, \ldots, \mathcal{X}^{OUT}_N\}$

A productivity shock ($\mathrm{d}\log \boldsymbol{A}$) and a markup shock ($\mathrm{d}\log \boldsymbol{\mu}$) effect in real GDP, Y can be decomposed into a pure change in technology ($\mathrm{d}\log \boldsymbol{A}$) for a given fixed $\mathcal{X}$ and the change in the distribution of $\mathcal{X}$ ($\mathrm{d}\mathcal{X}$) holding technology constant. In vector notation:

$$\mathrm{d}\log \mathcal{Y} = \underbrace{\frac{\partial \log \mathcal{Y}}{\partial \log A}\mathrm{d}\log A}_{\Delta \text{ Technology}} + \underbrace{\frac{\partial \log \mathcal{Y}}{\partial \mathcal{X}}\mathrm{d}\log \mathcal{X}}_{\Delta \text{ Allocative Efficiency}} \tag{3}$$

First, as a benchmark, we consider the case of a single-good firm without a joint production structure. The following gives a formal definition,

Definition 3. Single-good firm: A single-good firm is one that offers the same price to all buyers based on the same marginal cost and the same markup.

Under this single good firm assumption, the difference between each good disappears, and we obtain the formula introduced by Baqaee and Farhi (2020),

$$d\log TFP = \underbrace{\sum_{i\in N} \tilde{\lambda}_i d\log A_i}_{\Delta \text{ Technology}} - \underbrace{\sum_{i\in N} \tilde{\lambda}_i d\log \mu_i - \sum_{f\in F} \tilde{\Lambda}_f d\log \Lambda_f}_{\Delta\text{(Between firm) Allocative Efficiency}} \tag{4}$$

The impact of firm-level shocks on TFP can be summarized by multiplying the productivity shock to firm i by the (cost-based) Domar weight. Furthermore, allocative efficiency can be summarized by changes in a) markups and b) factor shares. If the initial equilibrium is inefficient, then the inefficient producer is too small relative to the optimal allocation to begin with. A decrease in the factor share represents a reallocation of resources to a more distorted part of the economy. On the other hand, if the change in factor share is driven by a change in markup, this does not imply a reallocation, and this must be discounted. Under the single good firm assumption, all misallocations arise between-firms, so we call the second term in equation (4) the between-firm misallocation.

We define within-firm AE as the difference between the allocative efficiency of equation (3) under the arbitrary CRS joint production production set and the allocative efficiency implied by the single good firm assumption from definition 3. The following proposition holds for the production set of CRS joint production

Proposition 1. *First-order approximation of within-firm AE:*

$$\Delta \textit{Within-Firm AE} = \sum_{i \in N} \tilde{\lambda}_i \underbrace{Cov_{s_i}\left(d \log \boldsymbol{p}_i, \frac{1}{\boldsymbol{\gamma}_i}\right)}_{\textit{Within-Firm AE of firm } i}$$

where $\boldsymbol{\gamma}_i = (\gamma_{i1}, ..., \gamma_{iG})$ *and* $d \log \boldsymbol{p}_i = (d \log p_{i1}, ..., d \log p_{iG})$.

This proposition provides sufficient statistics to give a first-order approximation of the contribution of within-firm AE to TFP. The contribution of each product to within-firm AE depends on the change in the price of the good relative to the firm's average price change and the distortion γ_{ig} of the relative network of that product in the firm's product portfolio. Thus, for example, if the price change for a given product g in firm i is higher than the average price change in firm i, and product g is more distorted relative to the average product within firm i, the covariance will be positive. Intuitively, if markups are accumulating in the downstream supply chain and the entire supply chain is underproducing, lower prices for upstream inputs can reduce distortions in that downstream supply chain and improve TFP.

Our approach detects within-firm misallocation in the sense of how much the single-good, single-markup assumption is violated in the data when product-level data are available. If the assumption is valid, our term becomes exactly zero. The assumption of the same marginal cost and the same markup is unlikely to hold in reality, but its quantitative importance is an empirical question. Since the misallocation of literature usually makes this assumption with or without networks, our decomposition is valuable in that it quantifies the degree to which the assumption is violated while isolating the impact of the existing misallocation literature.

Combining Proposition 1 with equation (4) provides a complete characterization of the change in TFP, including between firm AE and within-firm AE. It is important to note that this result is not unpacking Baqaee and Farhi (2020) growth accounting formula, but adding a new source of resource misallocations that equation (4) does not capture: within-firm.

Corollary 1. *First-order decomposition of TFP:*

$$d\log TFP = \underbrace{\sum_i \tilde{\lambda}_i d\log A_i}_{\text{Technology}} \underbrace{- \sum_i \tilde{\lambda}_i d\log \mu_i - \sum_f \tilde{\Lambda}_{(f)} d\log \Lambda_f}_{\text{Between-Firm AE}} + \underbrace{\sum_i \tilde{\lambda}_i Cov_{s_i}\left(d\log \boldsymbol{p}_i, \frac{1}{\boldsymbol{\gamma}_i}\right)}_{\text{Within-Firm AE}}$$

If firms are single-product firms, they charge a unique markup over the same marginal costs, so the within-firm AE converges to zero, and we are back to Baqaee and Farhi (2020). Finally, if prices equal marginal cost, there are no markups, and then the Between firm AE also converges to zero, where all TFP changes are due to technology converging to Hulten (1978).

Discussion: Our formula has two advantages over the decomposition formulas (4). First, our formula allows for a more general joint production structure. If we consider different products as different firms and apply the formula (4) to the same data, we need to assume separate productivity functions for each product among the same firms and the input distribution for each production function. Second, the allocative efficiencies calculated under the above assumptions cannot be decomposed into within-firm and between-firm.

Our formula can also be used to separate the within-firm AE resulting from any price heterogeneity within the same firm from the between-firm AE as long as the data are available. For example, if there is price variation for different buyers of the same product, this heterogeneity can be analyzed by considering them as different products.

2.4 Illustrative examples

We build examples assuming the simple economy to gain intuition on how the theory works. The first example includes only between-firm AE, while the second includes only within-firm AE. Although our formulas work for general joint production, we present an example with separable within-firm production functions to illustrate the difference in AE between within-firm and between-firm allocations.

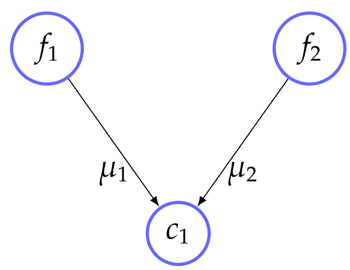
Throughout the three examples, there is a representative household with Cobb-Douglas utility.

$$U = c_1^{\alpha_1} c_2^{\alpha_2}$$

Note that by setting GDP to numeraire, the final expenditure share is determined from α only. We focus on the reallocation effect when firms have the same technology level A and only the markup changes.

Illustrative example 1: Between firm misalocation

Figure 1: Between firm misalocation



There are two single-product firms that produce using linear technology, and labor is inelastically supplied.

$$q_i = L_i, \quad \text{for } i = 1, 2$$

The allocation under the competitive setting is

$$\left(L_1^*, L_2^*\right) \propto \left(\frac{\alpha_1}{\mu_1}, \frac{\alpha_2}{\mu_2}\right)$$

While efficient allocation under marginal cost pricing is

$$\left(L_1^{SP}, L_2^{SP}\right) \propto (\alpha_1, \alpha_2)$$

Firm 1 is underproducing because it employs less labor relative to its optimal allocation than Firm 2.

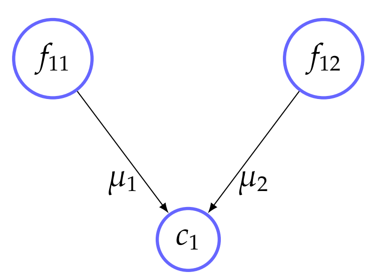
If there is a negative shock $d\log\mu_1 = -\epsilon$ to the markup of firm 1, this should lead to a reallocation of labor to the underproducing firm 1, improving AE.

$$
\begin{aligned}
d\log TFP &= \underbrace{-\alpha_1 d\log\mu_1 - d\log\Lambda_f}_{\text{Between Firm AE}} - \underbrace{\sum_i^2 \alpha_i Cov\left(d\log p_i, \frac{1}{\gamma_i}\right)}_{\text{Within Firm AE} = 0} \\
&= \alpha_1\epsilon - \frac{1}{1+\frac{\alpha_2\mu_1}{\alpha_1\mu_2}}\epsilon \\
&= \epsilon\alpha_1\left[1 - \frac{1}{\alpha_1 + \alpha_2\frac{\mu_1}{\mu_2}}\right] > 0
\end{aligned}
$$

First, the contribution of technology is 0 from the assumption. Second, Within Firm AE is zero from the definition of covariance since each firm produces only one good. The remaining firm misallocation is the labor share, which is $\Lambda_L = \left(\frac{\alpha_1}{\mu_1} + \frac{\alpha_2}{\mu_2}\right)$, and thus $d\log\Lambda_L = \frac{1}{1+\frac{\alpha_2\mu_1}{\alpha_1\mu_2}}\epsilon$.

Illustrative example 2: Within firm misallocation

Figure 2: Within firm misalocation



Next, we see how our formula works by setting the two products in the same setting and assuming that they are produced by the same firms. Of course, since we just relabeled the same economy, the reallocation effects for the same markup shocks should be the same.

However, the application of the formula changes as follows

$$d\log TFP = \underbrace{d\log A_i}_{\text{Technology = 0}} \underbrace{-d\log\mu - d\log\Lambda_f}_{\text{Between Firm AE=0}} - \underbrace{Cov_\alpha\left([\epsilon, 0], [\mu_1, \mu_2]\right)}_{\text{Within Firm AE}}$$
$$= \epsilon\alpha_1\left[1 - \frac{1}{\alpha_1 + \alpha_2\frac{\mu_1}{\mu_2}}\right] > 0$$

In this example, there is only one firm, so the between-firm allocation should be zero by definition. The change in TFP in Example 1 is calculated as a within-firm allocation in Example 2, yielding the same change in TFP. Thus, our formula correctly detects misallocations that occur within a firm.

3 Data

3.1 Data Sources

We need a granular product-level dataset to estimate TFP growth components, including product-level markups and product-level input-output matrices. While growth accounting in the presence of distortions can be performed using combinations of industry-level input-output tables and firm-level production variables as Baqaee and Farhi (2020) did, within industry production factor reallocations are omitted, which could mute part of the aggregate TFP movements. Moreover, when using firm-level data instead of product-level data, within-firm AE also disappears.

Leveraging on Chilean IRS sources, we use a decade of the universe of formal firms operating in Chile. Due to tax enforcement implications, every formal firm in Chile must declare all its invoices with other firms. The latter implies that for the universe of formal Chilean firms, we observe every product, quantity, and price traded with other firms. We also have access to tax accounting declarations from the Chilean IRS, which provides monthly data on each firm's revenue and input expenditures, including capital and labor.

We use data from four different sources of the Chilean IRS (Servicio de Impuestos Internos, SII). One of the advantages of SII data is that firms and workers have a unique identifier, which allows the merging of individuals and firms across data sets.

The first source used is the value-added tax form including gross monthly firm sales, materials expenditures, and investment. Second, the SII provides information from a matched employer-employee census of Chilean firms from 2005. Specifically, firms must report their employee's form that records all firms' payments to individual workers: the sum of taxable wages, overtime, bonuses, and any other labor earnings for each fiscal year. Since all legal firms must report to the SII, the data covers the total labor force with a formal wage contract, representing roughly 65% of employment in Chile[3]. For any given month, it is possible to identify the employment status of an individual worker, a measure of her average monthly labor income in that year, and a monthly measure of total employment and the distribution of average monthly earnings within the firm.

Third, data from the income tax form gathers yearly information on all sources of income and expenses of a firm. This form allows computing every individual's actual tax payments for each year. Even though details on sales and employment are available on this form, we use only data on capital stock for each firm and year to build perpetual inventories using data from the monthly F22 form. The user cost of capital is obtained by multiplying nominal capital stock by the real rental rate of capital. The real rental rate of capital is built using publicly available data. We use the 10-year government bond interest rate minus expected inflation plus the external financing premium. Also, we use the capital depreciation rate from the LA-Klems database.

Fourth, data from electronic tax documents (invoices universe) that provide information on each product, including its price and quantity, traded domestically or internationally with at least one Chilean firm participant from 2014.

There is an industry identifier for each firm at the 6-digit ISIC (rev. 4) level, allowing estimations from 9 (Chilean-specific industry classification levels) productive sectors up to more than 800 (when using six-digit sectors).

3.2 Multi-product firms data

We observe the universe of firm-to-firm transactions and firm's total sales (which include the latter and firm-to-consumer sales). We compute firm-specif product shares for the firm-to-firm universe and assume that its distributions are equivalent to firm-to-consumer transactions to recover the complete distribution of firm sales by product.

Every firm-to-firm transaction reports a "detail" column that records each product name transacted. "Details" are often firm idiosyncratic and differ from firm to firm for the same product[4]. We observe more than 15 million firm-"detail" pairs and keep track

[3]Central Bank of Chile (2018)

[4]For example, supermarket 1 declares selling "Sprite can 330cc", while supermarket 2 declares selling

of each pair's monthly quantity and price to generate appropriate price indices.

We aggregate products to a 290 product code identifier so they are comparable between firms. We do this to estimate product production functions that use the same product across firms. Table 2 shows the number of products distributed and sold by firms under the original "detail" product identifier and the 290 product code identifier.

Table 2: Number of products sold distribution

	"Detail" product	290 product code
Mean	41	32
Median	35	27
Sd	536	24
p1	10	6
p25	25	13
p75	102	47
p90	150	67
p95	364	78
p99	626	102
Max	92,142	181

3.3 Data Treatment

The data is anonymized to ensure confidentiality regarding the firm's and workers' identities. A set of filters is applied over the raw data to obtain the final data set for the empirical analysis. First, for the complete data set, a firm is defined as a taxpayer with a tax ID, positive sales, positive materials, positive wage bill, and capital for any given year. Second, firms that hire less than two employees or capital valued below US$20 a year are dropped. Third, all variables are winterized at 1% and 99% levels to avoid as much measurement error as possible. These criteria generate an economy-wide yearly firm panel for 2014-2022.

We start our data analysis by obtaining firm accounting variables from IRS registers. The level of capital is derived using the perpetual inventory method. At the same time, worker headcounts are directly observed from the data. Similarly, the IRS directly collects quantities produced and summarizes the intermediate input. However, due to our product aggregation (from around 15 million products to 290 product codes), it is necessary to create aggregated product-level quantity produced and material usage indices.

"Sprite sugar beverage 330".

We build each firm's product code-level output and intermediate goods input price indices using standard Tornqvist indices. We selected 2014 as the base year for constructing our price indices because it was the first year in which we observed prices for firm-to-firm transactions. This method is widely recognized for estimating aggregate production functions at the firm or plant level when price data is accessible (Dhyne et al. (2022) and De Roux et al. (2021)).

3.4 Discussion on measurements

Before starting the applications, we will discuss the data implementation, which is not straightforward when aggregated in inefficient economies with multi-product firms. Baqaee and Farhi (2020) faced cost, production factors observability, and firm-level data issues, which, due to the granular database we use, are not binding for us; from tax records, we directly observe firm-level cost data and production factor usage. Nevertheless, we face challenges in mapping the model wedges, which are product-level, to the data.

To construct product-level network distortions, we need information on the ratio of product-level sales to cost-based Dormar weights and markups. The two former can be observed directly from our data without imputation, which captures the network distortion to which the product is exposed. On the other hand, the distortion of the product itself, i.e., the markup, needs to be estimated as in the rest of the literature.

However, we cannot use the literature workhorse markup estimation strategy, De Loecker and Warzynski (2012). While we could use De Loecker et al. (2016) strategy on multi-product firms markups, that work relies on non-joint production setups, where each product production process is independent of other products within a firm. The latter approach is not that appealing to us because we want to capture the inefficiencies of the production factors within a firm when the production changes of one product affect the production of the other products in the firm, which have evidence support based on Ding (2023).

To recover product level cost information and product level markups, we rely on the seminal work by Dhyne et al. (2022), which developed an estimation method for multi-product production functions assuming joint production.

Our benchmark estimation allows firms to produce at most 5 of the 290 available product codes. We restrict product codes to account for at least 20% of the firm's total sales. All other goods that represent less than 20% of the firm's total sales are grouped into a composite good that combines all the remaining products[5].

4 Application: Growth accounting in the presence of multi-product markups

We bring our theory to the data to test the importance of within-firm allocative efficiency. As established in Section 2, to compute within and between allocative efficiency, one needs to measure four objects: (1) product level markups , (2) cost-based Domar weights $\tilde{\lambda}$, (3) product distortion in networks and (4) aggregates such as value-added factor changes. We discuss each of these in turn.

4.1 Markups

Based on Dhyne et al. (2022), we perform joint production function estimations, computing product-specific markups. In a joint production setup, firms use common inputs to produce a product portfolio, meaning that some inputs may simultaneously be used to produce multiple products. They proposed an Ackerberg et al. (2015) like production function estimation method based on Diewert's (1973) production set. The following is an overview of Dhyne et al. (2022)'s methodology.

A firm has production possibilities set, V, that consists of a set of feasible inputs $x = (x_1, ..., x_M)$ and outputs of the product, $q = (q_1, ..., q_G)$. For any (q_g, x) the transformation function is defined as

$$q_g^* = f_g\left(q_g, x\right) \equiv \max\left\{q_g | \left(q_g, q_{-g}, x\right) \in V\right\}$$

To identify the unobserved marginal cost for each firm's product, we rely on (variable) cost minimization. Firms have $N-1$ freely variable inputs and one fixed input, capital (K), so the problem that a firm faces to minimize its variables cost to produce its output vector q_i^* given the input prices vector $p_x = (p_{x1}, ..., p_{xM})$ and unobserved productivity for products, $\omega = (\omega_1, ..., \omega_G)$.

Defining the Lagrangian multiplier of the cost minimization problem, λ_g, as the marginal

[5]We keep a record of product-specific prices and quantities to build price indices of the composite good.

cost of product g, the first order condition for every optimal input demand yield:

$$p_m = \lambda_g \frac{\partial f(q^*_{-g}, x, K, \omega)}{\partial x_m} \;\; \forall m = 1, .., M,$$

It is possible to solve for product g marginal cost as the expenditure on production input m divided by its output elasticity (β^g_n) times product g production:

$$\lambda_g = \frac{p_m}{\frac{\partial f(q^*_{-g}, x, K, \omega)}{\partial x_m}} = \frac{p_m x^*_m}{\beta^g_m q^*_g},$$

Multiplying the marginal cost expression by $\frac{1}{p_g}$, where p_g is product g price, product g markup is given by:

$$\mu_g = \beta^g_m \, \frac{p_g \, q^*_g}{p_m \, x^*_m},$$

Markup identification is reached following Dhyne et al. (2022), built in the standard single product markup literature. We use control functions for the unobserved productivity terms (i.e., Ackerberg et al. (2015)) to account for unobserved productivity with the difference of the need for instruments for q_{-g}; following Dhyne et al. (2022) we use lagged values of q_{-g}.

We assume that firms use a Cobb-Douglas production function with three factors: (Capital k, Labor l, and Materials m). A multi-product firm will produce physical units of product g using the following production function:

$$\log q_{gt} = \beta^g_0 + \beta^g_k \log k_t + \beta^g_l \log l_t + \beta^g_m \log m^j_t + \gamma^g_{-g} \; \log q_{-gt} + \omega_{gt}$$

We pool together products at one digit (9 aggregate product categories) and perform production function estimations separately for each category following ACF using a GMM estimator. Table 3 shows the estimation coefficients. The materials coefficient is significant for every product category, and the other goods coefficient (γ_{-g}) is always negative and significant, implying that all marginal costs are positive.

Table 3: Production function estimation coefficients by product category

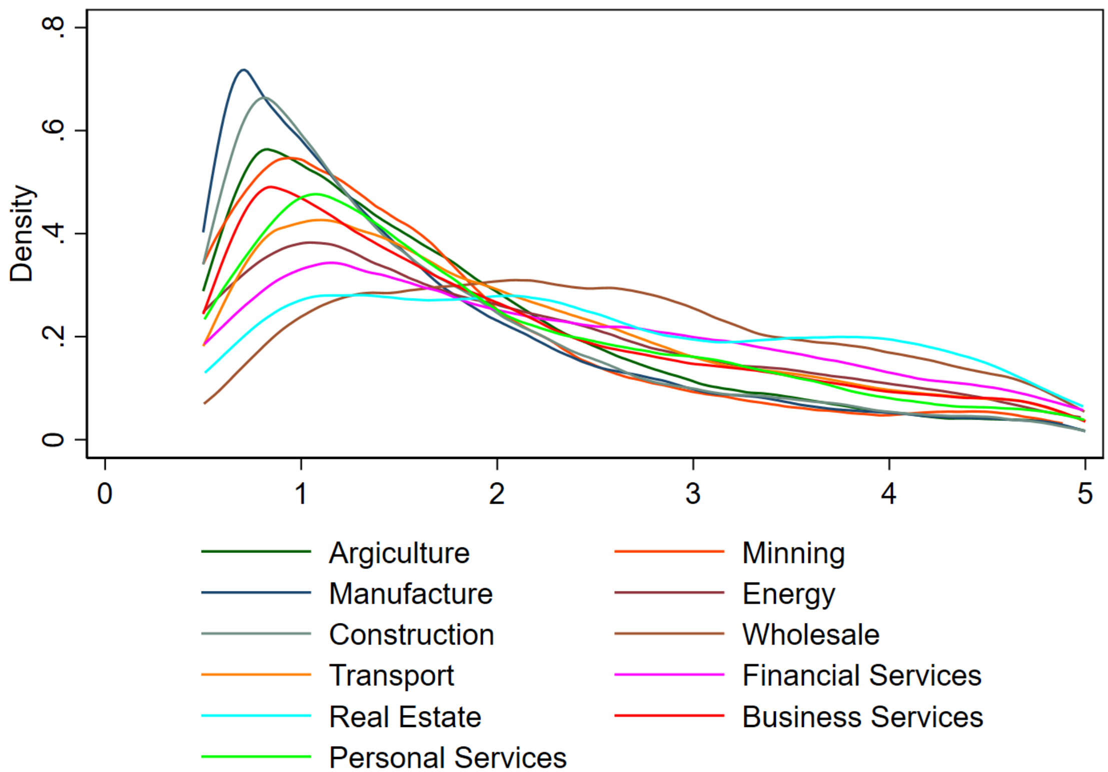
	β_m	β_l	β_k	γ_{-g}	N obs
Agriculture products	0.77***	0.19***	0.06**	-0.13***	183,465
Mining products	0.59***	0.19***	0.15**	-0.09***	38,958
Manufacturing products	0.86***	0.55***	0.07***	-0.44***	343,539
Energy products	0.80***	0.47***	0.04	-0.31***	7,938
Construction products	0.67***	0.43***	0.03*	-0.25***	151,909
Commerce products	1.14***	0.23***	0.04***	-0.26***	429,423
Transport and ICTs products	0.50***	1.11***	0.03**	-0.08***	161,151
Financial and real estate Activities products	0.71***	0.52***	0.04**	-0.24***	49,968
Other services products	0.72***	0.70***	0.02**	-0.38***	239,176

Note: *** p<0.01, ** p<0.05, * p<0.1

Product-level markup distribution concentrated around 1, with a 1.22 median (Figure 3). We remain agnostic about product-level markup interpretation. While the markup distribution offers valuable insights into diagnosing the presence of product market power, it can potentially lead to misleading conclusions regarding allocative efficiency. Markup distribution does not account for the significance of firms within production networks.

As noted above, we calculate the allocative efficiency resulting from the above markups by using an approach that captures the inefficiencies arising from the linkages between networks and between and within firms.

Figure 3: Product level markup distribution in 2018



Notes: The figure shows .

4.2 Cost-based Domar weights

The estimation procedure for the Corollary 1 equation, which includes Proposition 1, can be summarized in two steps. In the first step, we utilize data on all firm-to-firm transactions and factor expenditures to construct, on an annual basis, each element of the cost-based input-output matrix by product denoted as $\tilde{\Omega}$. Specifically, we compute the denominator of each element (indexed by ig, jp) by summing a firm's purchases from all its suppliers, its wage bill, and its capital level multiplied by its relevant user cost rental rate of capital. The last two elements of the matrix have wage bills and capital expenditures as their numerators. Therefore, $\tilde{\Omega}$ is $(DG+F) \times (DG+F)$ matrix, which is composed of following elements:

$$\tilde{\Omega} = \left[\begin{array}{c|ccc|cc} 0 & \tilde{\Omega}_{C,11} & \cdots & \tilde{\Omega}_{CN} & 0 & 0 \\ \hline 0 & \tilde{\Omega}_{11,11} & \cdots & \tilde{\Omega}_{11,NG} & \tilde{\Omega}_{11L} & \tilde{\Omega}_{11K} \\ 0 & & \ddots & & & \\ 0 & \tilde{\Omega}_{NG,11} & & \tilde{\Omega}_{NG,NG} & \tilde{\Omega}_{NGL} & \tilde{\Omega}_{NGK} \\ \hline 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{array}\right].$$

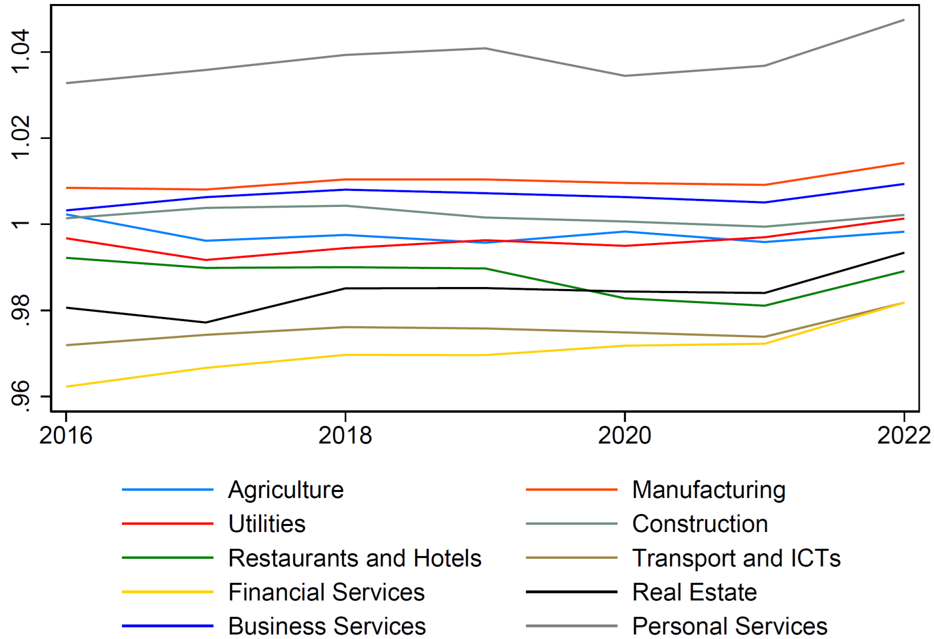
In the second step, we proceed to derive Cost-Based Domar weights. This involves two sub-components. The first sub-component is the cost-based Leontief inverse ($\tilde{\Psi}$). The second sub-component comprises the b vector. Each element of this vector represents the final consumption of firms and is computed by subtracting intermediate sales (sales to other firms recorded in the revenue-based IO matrix) from a firm's total sales. Combining both sub-components, we compute the product-level cost-based Domar weight ($\tilde{\Lambda}$) by multiplying the transpose of the b vector with the cost-based Leontief inverse ($\tilde{\Psi}$). The technology term is then built as a residual by subtracting Between and within firm AE terms from TFP growth.

4.3 Product distortion in networks

We then compute product-level network distortions by dividing each product's total cost (which we compute by dividing the product's total sales by its markup) normalized by GDP by its correspondent cost-based Domar weight.

Figure 4 shows the median of product distortion by product categories relative to the manufacturing product. All industries show different median product distortions evolution. Moreover, their level is consistently different, which implies that the other products have different network distortions on average. For example, distortions in the mining product (orange) are smaller than those in the manufacturing industry (green), but firms producing goods in these two categories are likely to underproduce manufacturing goods on average.

Figure 4: median of Γ_{ig} by year

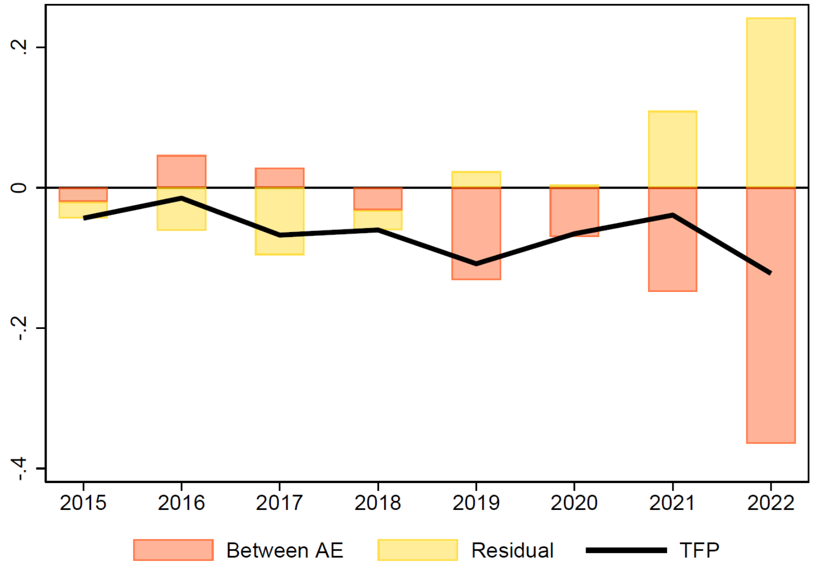


Notes: This Figure shows the change over time in the value of the median in the twelve major categories of products, normalized to 1 for the mining product.

4.4 Aggregate TFP growth

We start with the single-good firm assumption. This means that, like Baqaee and Farhi (2020), we only consider between-firm misallocation. Figure 6 decomposes this into the cumulative change in TFP from 2014 to 2022. The between-firm misallocation (red) overshoots, particularly during COVID-19 and the subsequent period of high inflation, while the technology factor (yellow), calculated as the residual, makes a significant positive contribution. Overall, between-firm misallocation by itself cannot fully explain the dynamics of TFP.

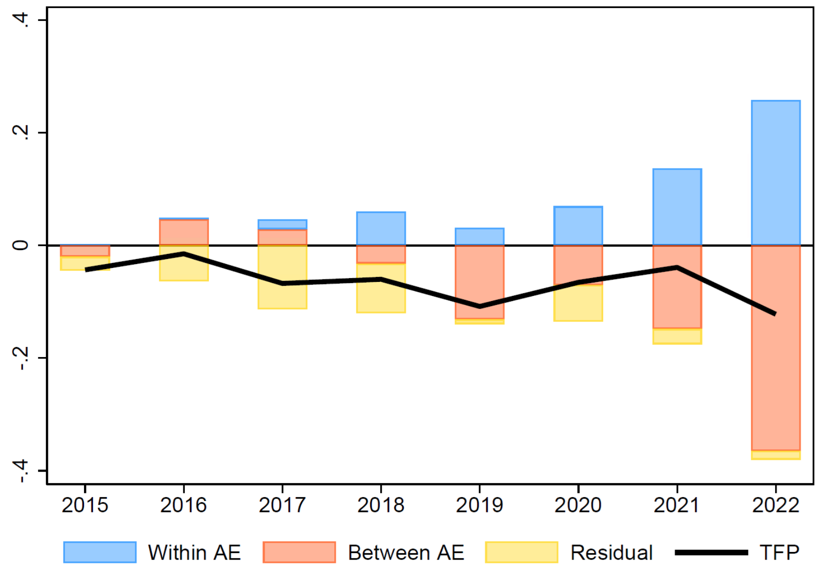
Figure 5: TFP decomposition with between-firm allocative efficiency



Notes: This Figure shows the cumulative change calculated by applying equation (4) repeatedly each year starting in 2014. Technology (residual) is calculated by subtracting between-firm allocative efficiency from TFP.

Next, we add sufficient statistics representing within-firm misallocation. Note that this section does not decompose between-firm misallocation but adds a new margin. Within-firm misallocation makes a smaller contribution in normal times before 2019 than between-firm misallocation. However, after 2020, it made a large upward contribution to COVID-19 and the subsequent period of high inflation, which reduced many of the residuals seen in Figure 5. In other words, within-firm misallocation, together with between-firm misallocation, accounts for TFP movements during COVID-19 and the high inflation period well.

Figure 6: TFP decomposition with between and within-firm allocative efficiency

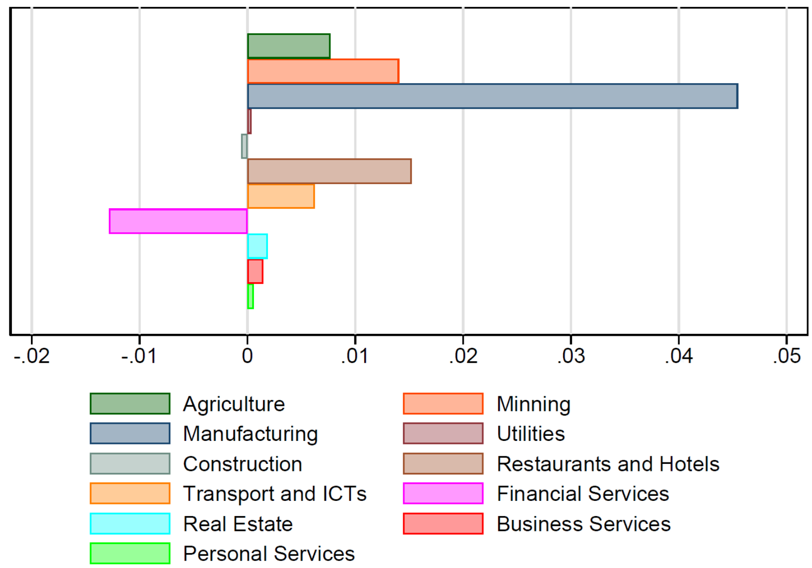


Notes: This Figure shows the cumulative change calculated by applying equation from Corollary (1) repeatedly each year starting in 2014. Technology (residual) is calculated by subtracting between and within-firm allocative efficiency from TFP.

To explore this further, we decompose the cumulative changes of within-firm allocative efficiency by broad product categories. We categorize firms based on the broad category of their main product (using the product with the biggest sales within a given firm) and plot the cumulative changes from 2019 to 2022 for each category. So, aggregate contribution by product categories of the firm's main product is

$$\text{Contribution of firm with product } c = \sum_{i \in N_C} \tilde{\lambda}_i Cov_{s_i}\left(d \log \boldsymbol{p}_i, \frac{1}{\gamma_i}\right)$$

Figure 7: TFP changes by categories from 2019 to 2022



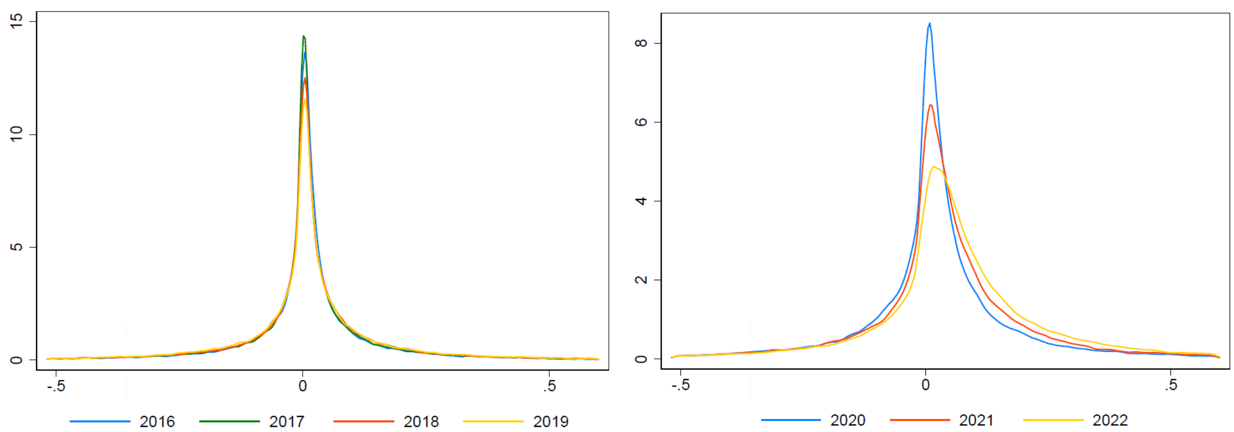
Notes: This Figure shows the cumulative log change calculated by repeatedly applying the equation from Corollary (1) each year starting in 2014. Technology (residual) is calculated by subtracting between-firm allocative efficiency from TFP.

The largest contribution is in the manufacturing product category, but except for construction and finance, generally, all goods make a positive contribution. Returning to the definition of covariance representing within-firm misallocation, this suggests that the prices of goods that were originally highly distorted tended to fall during the crisis. This implies that, on average, the within-firm allocative efficiency improved during the crisis by shifting the firm's main product from the highly distorted good to other less distorted goods that were typically sold less heavily.

The granularity of the data also makes it possible to track changes over time in the distribution of allocative efficiency (i.e., covariance) within firms. Since covariance degenerates to zero when the single-good firm assumption holds, the fact that covariance is dispersed implies that within-firm misallocation exists and varies widely across firms. We find that the distributions vary significantly from period to period. First, the left panel 8 plots the distribution for normal times (pre-COVID-19, 2016-2019), which is symmetric around 0, with small differences from year to year. This results in a smaller overall contribution (see contributions through 2019 in Figure 6).

Figure 8: $Cov_{s_i}\left(d\log \boldsymbol{p}_i, \frac{1}{\gamma_i}\right)$ distributions by year

A. Normal periods (2016 to 2019) B. COVID-19 and high-inflation periods



Notes: These Figures plot the distribution of firm-level $Cov_{s_i}(d\log \boldsymbol{p}_i, \xi_i)$ for each year. The within-firm misallocation in Proposition 1 and Corollary 1 is recovered by weighting this distribution with firm-level cost-based Dormer weights.

On the other hand, the right Figure 8 plots the distribution after COVID-19. The distribution shifts to the right from year to year, resulting in a right-skewed distribution. This suggests that the increase in within-firm allocative efficiency is not being driven by a few specific firms but rather by a broad range of firms that are reallocating to less distorted products within the firm. This is consistent with the fact that many product categories contribute to within-firm allocative efficiency in product-by-product decomposition. Overall, our results suggest that product-level information is important for the analysis of misallocation.

5 Conclusion

We build a theory to include within-firm AE and evaluate its contribution to aggregate TFP growth. Relying on a granular database for Chile, including each product, price, and quantity transacted between two formal firms in the economy, we find that within-firm AE (how firms choose their product portfolios) accounts for around one-fourth of TFP growth in the last decade in Chile.

We look forward to extending our work to use between and within-firm AE measures to identify the relatively more distorted products, firms, and industries where industrial policies could be more effective. The aim of the extension is to understand the mechanisms through which industrial policy can reduce product-specific distortions and evaluate the aggregate TFP implications of targeted industrial policies through counterfactuals using models.

References

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Baqaee, D., Burstein, A., Duprez, C., and Farhi, E. (2023). *Supplier Churn and Growth: A Micro-to-Macro Analysis*. National Bureau of Economic Research.

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

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

Bigio, S. and La'O, J. (2020). Distortions in production networks*. *The quarterly journal of economics*, 135(4):2187–2253.

Cairncross, J. and Morrow, P. (2023). Multi-Product markups.

Central Bank of Chile (2018). Mercado laboral: hechos estilizados e implicancias macroeconómicas. *Central Bank of Chile Publications*.

Dávila, E. and Schaab, A. (2023). Welfare accounting.

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., and Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica: journal of the Econometric Society*, 84(2):445–510.

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American economic review*, 102(6):2437–71.

De Roux, N., Eslava, M., Franco, S., and Verhoogen, E. (2021). Estimating production functions in differentiated-product industries with quantity information and external instruments. Technical report, National Bureau of Economic Research.

Dhyne, E., Petrin, A., Smeets, V., and Warzynski, F. (2017). Multi product firms, import competition, and the evolution of firm-product technical efficiencies.

Dhyne, E., Petrin, A., Smeets, V., and Warzynski, F. (2022). Theory for extending single-product production function estimation to multi-product settings. Technical report, National Bureau of Economic Research.

Diewert, W. E. (1973). Functional forms for profit and transformation functions. *Journal of Economic theory*, 6(3):284–316.

Ding, X. (2023). *Industry Linkages from Joint Production*. U.S. Census Bureau, Center for Economic Studies.

Hall, R. (1988). The relation between price and marginal cost in U.S. industry. *The journal of political economy*, 96(5):921–947.

Hall, R. E. (1973). The specification of technology with several kinds of output. *The journal of political economy*, 81(4):878–892.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448.

Hulten, C. R. (1978). Growth accounting with intermediate inputs. *The Review of Economic Studies*, 45(3):511–518.

Kikkawa, A. K. (2022). Imperfect competition in Firm-to-Firm trade. *Journal of the European Economic Association*, 20(5):1933–1970.

Liu, E. (2019). Industrial policies in production networks*. *The quarterly journal of economics*, 134(4):1883–1948.

Osotimehin, S. and Popov, L. (2023). Misallocation and intersectoral linkages. *Review of economic dynamics*.

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

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

Solow, R. M. (1957). Technical change and the aggregate production function.

Valmari, N. (2023). Estimating production functions of multiproduct firms. *The Review of economic studies*.

Appendix

A Proof

Assume the implicit function theorem can be applied so that we could pick some reference good r of firm i and explicitly functional form locally: $y_{ir} = f_{ir}\left(\boldsymbol{y}_{i,-r}, \boldsymbol{x}_i, \boldsymbol{L}_i\right)$ where -r refer all $g \neq r.\boldsymbol{x}_i$ is the vector of intermediate goods and $\boldsymbol{L}_i$ is the vector of factors. Then, we could solve the cost-minimization problem:

$$C_i\left(\boldsymbol{q}_i, \boldsymbol{p}_i\right) = \sum_f w_{if} x_f + mc_{ir}\left(y_{ir} - f_{ir}\left(\boldsymbol{y}_{i,-r}, \boldsymbol{x}_i, \boldsymbol{L}_i\right)\right)$$

By taking a derivative with respect to y_{ig} for $g \neq r$, we have

$$mc_{ig} = mc_{ir}\left(-\frac{\partial f_{ir}\left(\boldsymbol{y}_{i,-r}, \boldsymbol{x}_i, \boldsymbol{L}_i\right)}{\partial y_{ig}}\right)$$

First, we introduce the following lemma.

Lemma 1. *Total derivative of* $mc_{ig} = mc_{ir}\left(-\frac{\partial f_{ir}\left(\boldsymbol{y}_{i,-r}, \boldsymbol{x}_i, \boldsymbol{L}_i\right)}{\partial y_{ig}}\right)$

$$\begin{aligned} d\log mc_{ig} =& d\log mc_{ig} + d\log y_{ig} - d\log y_{ir} + \sum_k \left(-\frac{\partial \log f_{ir}\left(y_{i,-r}, x\right)}{\partial \log y_{ig} \partial \log y_k}\right) d\log y_k + \\ &+ \sum_{j,p}\left(-\frac{\partial \log f_i\left(y_{-g}, \boldsymbol{x}_i\right)}{\partial \log y_{ig} \partial \log x_{i,jp}}\right) d\log x_{i,jp} + \sum_f \left(-\frac{\partial \log f_i\left(y_{-g}, \boldsymbol{x}_i\right)}{\partial \log y_{ig} \partial \log L_{if}}\right) d\log L_{if} \end{aligned}$$

Proof. First order approximation of $mc_{ig} = mc_{ir}\left(-\frac{\partial f_{ir}\left(\boldsymbol{y}_{i,-r}, \boldsymbol{x}_i, \boldsymbol{L}_i\right)}{\partial y_{ig}}\right)$

$$\begin{aligned} d\log mc_{ig} =& d\log mc_{ig} + d\log y_{ig} - d\log y_{ir} + d\log\left\{-\frac{\partial \log f_{ir}\left(y_{i-r}, x\right)}{\partial \log y_{ig}}\right\} \\ =& d\log mc_{ig} + d\log y_{ig} - d\log y_{ir} + \sum_k \left(-\frac{\partial \log f_{ir}\left(y_{i,-r}, x\right)}{\partial \log y_{ig} \partial \log y_k}\right) d\log y_k \\ &+ \sum_{j,p}\left(-\frac{\partial \log f_i\left(y_{-g}, \boldsymbol{x}_i\right)}{\partial \log y_{ig} \partial \log x_{i,jp}}\right) d\log x_{i,jp} + \sum_f \left(-\frac{\partial \log f_i\left(y_{-g}, \boldsymbol{x}_i\right)}{\partial \log y_{ig} \partial \log L_{if}}\right) d\log L_{if} \end{aligned}$$

□

Next, we proof proposition 1

Proof. By CRS, we know

$$C_i(\boldsymbol{q}_i, \boldsymbol{p}_i) = \sum_g q_{ig} mc_{ig}$$

Total derivative:

$$\begin{aligned}
RHS &= \sum_g \frac{\partial \log\left(\sum q_{ig} mc_{ig}\right)}{\partial \log y_i} y_i d\log y_i + \sum_i \frac{\partial \log\left(\sum q_{ig} mc_{ig}\right)}{\partial \log mc_{ig}} mc_{ig} d\log mc_{ig} \\
&= \sum_g \frac{q_{ig} mc_{ig}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log q_{ig} + \sum_g \frac{q_{ig} mc_{ig}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log mc_{ig}
\end{aligned}$$

and

$$LHS = -d\log A_i + \sum_{j,p} \frac{p_{jp} x_{i,jp}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log p_{jp} + \sum_f \frac{w_f l_{i,f}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log w_f + \sum_g \frac{q_{ig} mc_{ig}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log q_{ig}$$

Hence,

$$\sum_g \frac{y_{ig} mc_{ig}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log mc_{ig} = -d\log A_i + \sum_{j,p} \frac{p_{jp} x_{i,jp}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log p_{jp} + \sum_f \frac{w_f L_{if}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log w_f \tag{5}$$

Rearranging this expression by picking one reference good, r

$$\begin{aligned}
\frac{q_{ir} mc_{ir}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log mc_{ir} &= -d\log A_i + \sum_{j,p} \frac{p_{jp} x_{jp}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log p_{jp} + \sum_f \frac{w_f L_{if}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log w_f + \sum_{g\neq r} \left(-\frac{q_{ig} mc_{ig}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)}\right) d\log mc_{ig} \iff \\
\frac{q_{ir} mc_{ir}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log p_{ir} &= -d\log A_i + \sum \frac{q_{ig} mc_{ig}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log \mu_{ig} + \sum_{j,p} \frac{p_{jp} x_{i,jp}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log p_{jp} + \sum_f \frac{w_f L_{if}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log w_f \\
&\quad + \sum_{g\neq r} \left(-\frac{q_{ig} mc_{ig}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)}\right) d\log p_{ig}
\end{aligned}$$

Let

$$d\log p_{ig}/p_{ir} = \Theta_{ig} \tag{6}$$

Combining 5 with 6 yields

$$\frac{q_{ir} mc_{ir}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log p_{ir} = -d\log A_i + \sum_g \frac{q_{ig} mc_{ig}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log \mu_{ig} + \sum_{j,p} \frac{p_{jp} x_{jp}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log p_{jp} + \sum_f \frac{w_f l_{if}}{C_i(\boldsymbol{q}_i, \boldsymbol{p}_i)} d\log w_f$$

$$+\sum\left(-\frac{q_{ig}mc_{ig}}{C_i\left(\boldsymbol{q}_i,\boldsymbol{p}_i\right)}\right)d\log p_{ig} \iff$$

$$\frac{q_{ir}mc_{ir}}{C_i\left(\boldsymbol{q}_i,\boldsymbol{p}_i\right)}d\log p_{ir} = -d\log A_i + d\log\mu_i + \sum_{j,p}\tilde{\Omega}_{ig,jp}d\log p_{jp} + \sum_f \tilde{\Omega}_{ig,f}d\log w_f + \sum_{g\neq r}\left(-\frac{q_{ig}mc_{ig}}{C_i\left(\boldsymbol{q}_i,\boldsymbol{p}_i\right)}\right)\left[d\log p_{ir}+\Theta_{ig}\right] \iff$$

$$d\log p_{ir} = -d\log A_i + d\log\mu_i + \sum_{j,p}\tilde{\Omega}_{ig,jp}d\log p_{jp} + \sum_f \tilde{\Omega}_{ig,f}d\log w_f + \sum_{g\neq r}\left(-\frac{q_{ig}mc_{ig}}{C_i\left(\boldsymbol{q}_i,\boldsymbol{p}_i\right)}\right)\Theta_{ig}$$

Since $\Theta_{ig} = 0$ if g is a reference good, the price equation could be expressed by

$$d\log p_{ig} = -d\log A_i + d\log\mu_i + \sum_{j,p}\tilde{\Omega}_{ig,jp}d\log p_{jp} + \sum_f \tilde{\Omega}_{ig,f}d\log w_f + \left\{\mathbb{I}_i(g) - \sum_{g\neq r}\left(\frac{q_{ig}mc_{ig}}{C_i\left(\boldsymbol{q}_i,\boldsymbol{p}_i\right)}\right)\right\}\Theta_{ig}$$

In vector notation

$$d\log\boldsymbol{p} = (I-\tilde{\Omega}^{NG\times NG})^{-1}\left\{-d\log\boldsymbol{A}^{NG\times 1} + d\log\boldsymbol{\mu}^{NG\times 1} + \tilde{\Omega}_f^{NG\times F}d\log\boldsymbol{w} + (\mathbf{1}-\mathbf{C})\circ\boldsymbol{\Theta}^{NG\times 1}\right\},$$

where $\circ$ represents Hadamard product and and C is a vector of $NG\times 1$, with the following C_i common elements for firm $i\in N$,

$$C_i = \sum_{g\neq r}\left(\frac{q_{ig}mc_{ig}}{C_i\left(\boldsymbol{q}_i,\boldsymbol{p}_i\right)}\right),$$

We know

$$d\log Y = -\boldsymbol{b}'d\log\boldsymbol{p},$$

Therefore, we have

$$\begin{aligned} d\log Y &= -\boldsymbol{b}'\tilde{\Psi}^{NG\times NG}\left\{-d\log\boldsymbol{A} + d\log\boldsymbol{\mu} + \tilde{\Omega}_f d\log\boldsymbol{w} + (\mathbf{1}-\mathbf{C})\circ\boldsymbol{\Theta}^{NG\times 1}\right\} \iff \\ &= -\tilde{\boldsymbol{\lambda}}'\left\{-d\log\boldsymbol{A} + d\log\boldsymbol{\mu} + \tilde{\Omega}_f d\log\boldsymbol{w} + (\mathbf{1}-\mathbf{C})\circ\boldsymbol{\Theta}^{NG\times 1}\right\}\end{aligned}$$

subtracting $\sum_f \tilde{\Lambda}_{(f)}d\log L_f$ from both sides yields

$$\begin{aligned} d\log TFP = &\sum_i \tilde{\lambda}_i d\log A_i - \sum_i \tilde{\lambda}_i d\log\mu_i - \sum_f \tilde{\Lambda}_{(f)}d\log\Lambda_f \\ &- \sum_i\left(\sum_g \tilde{\lambda}_{ig}d\log p_{ig}/p_{ir} - \sum_{g\neq r}\frac{q_{ig}mc_{ig}}{C_i\left(\boldsymbol{q}_i,\boldsymbol{p}_i\right)}\tilde{\lambda}_i d\log p_{ig}/p_{ir}\right)\end{aligned}$$

Note that

$$
\begin{aligned}
\left(\sum_g \tilde{\lambda}_{ig} d\log p_{ig}/p_{ir} - \sum_{g\neq r} \frac{q_{ig} mc_{ig}}{C_i\left(\boldsymbol{q}_i, \boldsymbol{p}_i\right)} \tilde{\lambda}_i d\log p_{ig}/p_{ir}\right) &= \tilde{\lambda}_i \left(\sum_g s_{ig} d\log p_{ig}/p_{ir} - \sum_{g\neq r} c_{ig} d\log p_{ig}/p_{ir}\right) \\
&= \tilde{\lambda}_i \left(\sum_g s_{ig} d\log p_{ig}/p_{ir} - \sum_{g} c_{ig} d\log p_{ig}/p_{ir}\right) \\
&= \tilde{\lambda}_i \left(\sum_g \left(s_{ig} - c_{ig}\right) d\log p_{ig}\right) \\
&= \tilde{\lambda}_i \left(\sum_g \left(s_{ig} - \frac{\frac{q_{ig} mc_{ig}}{C\left(y,p\right)}}{s_{ig}} s_{ig}\right) d\log p_{ig}\right) \\
&= \tilde{\lambda}_i \left(\sum_g \left(s_{ig} - \xi_{ig} s_{ig}\right) d\log p_{ig}\right) \\
&= \tilde{\lambda}_i \left(E_{s_i}\left[d\log \boldsymbol{p}_i\right] E_{s_i}\left[\xi_i\right] - E_{s_i}\left[d\log \boldsymbol{p}_i, \boldsymbol{\xi}_i\right]\right) \\
&= -\tilde{\lambda}_i Cov_{s_i}\left(d\log \mathbf{p}_i, \boldsymbol{\xi}_i\right)
\end{aligned}
$$

where $c_{ig} = \frac{q_{ig} mc_{ig}}{C_i\left(\boldsymbol{q}_i, \boldsymbol{p}_i\right)}$. □