The Microeconomic Foundations of Aggregate Production Functions

David Rezza Baqaee Emmanuel Farhi*
UCLA Harvard

July 22, 2019

Abstract

Aggregate production functions are reduced-form relationships that emerge endogenously from input-output interactions between heterogeneous producers and factors in general equilibrium. We provide a general methodology for analyzing such aggregate production functions by deriving their first- and second-order properties. Our aggregation formulas provide non-parameteric characterizations of the macro elasticities of substitution between factors and of the macro bias of technical change in terms of micro sufficient statistics. They allow us to generalize existing aggregation theorems and to derive new ones. We relate our results to the famous Cambridge-Cambridge controversy.

*Emails: baqaee@econ.ucla.edu, efarhi@harvard.edu. We thank Maria Voronina for excellent research assistance. We thank Natalie Bau and Elhanan Helpman for useful conversations.

1 Introduction

The aggregate production function is pervasive in macroeconomics. The vast majority of macroeconomic models postulate that real GDP or aggregate output Y can be written as arising from some specific parametric function $Y = F(L_1, \ldots, L_N, A)$, where L_i is a primary factor input and A indexes different production technologies. By far the most common variant takes the form $Y = AF(A_K K, A_L L)$, where A, A_K, and A_L index Hicks-neutral, capital-augmenting, and labor-augmenting technical change, and F is a CES function.[1]

From the early 50s to the late 60s, the aggregate production function became a central focus of a dispute commonly called the Cambridge-Cambridge controversy. The attackers were the post-Keynesians, based primarily in and associated with Cambridge, England, and the defenders were the neoclassicals, based primarily in and associated with Cambridge, Massachusetts.[2] A primary point of contention surrounded the validity of the neoclassical aggregate production function. To modern economists, the archetypal example of the neoclassical approach is Solow's famous growth model (Solow, 1956), which uses an aggregate production function with capital and labor to model the process of economic growth.

The debate kicked off with Joan Robinson's 1953 paper criticizing the aggregate production function as a "powerful tool of miseducation." The post-Keynesians (Robinson, Sraffa, and Pasinetti, among others) criticized the aggregate production function, and specifically, the aggregation of the capital stock into a single index number. They were met in opposition by the neoclassicals (Solow, Samuelson, Hahn, among others) who rallied in defense of the aggregate production function.

Eventually, the English Cambridge prevailed against the American Cambridge, decisively showing that aggregate production functions with an aggregate capital stock do not always exist. They did this through a series of ingenious, though perhaps exotic looking, "re-switching" examples. These examples demonstrated that at the macro level,

[1] More precisely, this variant can be written as

$$\frac{Y}{\bar{Y}} = \frac{A}{\bar{A}} \left(\omega_K \left(\frac{A_K K}{\bar{A}_K \bar{K}} \right)^{\frac{\sigma-1}{\sigma}} + \omega_L \left(\frac{A_L L}{\bar{A}_L \bar{L}} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}},$$

where bars denote values of output, factors, and productivity shifters, at a specific point, ω_L and $\omega_K = 1 - \omega_L$ are the and labor and capital shares at that point, and σ is the elasticity of substitution between capital and labor. The popular Cobb-Douglas specification obtains in the limit $\sigma \to 1$.

[2] See Cohen and Harcourt (2003) for a retrospective account of the controversy.

"fundamental laws" such as diminishing returns may not hold for the aggregate capital stock, even if, at the micro level, there are diminishing returns for every capital good. This means that a neoclassical aggregate production function could not be used to study the distribution of income in such economies.

However, despite winning the battle, the English side arguably lost the war. Although exposed as a fiction, the "neoclassical" approach to modeling the production technology of an economy was nevertheless very useful. It was adopted and built upon by the real business cycle and growth literatures starting in the 1980s. Reports of the death of the aggregate production function turned out to be greatly exaggerated, as nearly all workhorse macroeconomic models now postulate an exogenous aggregate production function.

Why did Robinson and Sraffa fail to convince macroeconomists to abandon aggregate production functions? One answer is the old adage: you need a model to beat a model. Once we abandon the aggregate production function, we need something to replace it with. Although the post-Keynesians were effective in dismantling this concept, they were not able to offer a preferable alternative. For his part, Sraffa advocated a disaggregated approach, one which took seriously "the production of commodities by means of commodities" (the title of his magnum opus). However, his impact was limited. Clean theoretical results were hard to come by and conditions under which factors of production could be aggregated were hopelessly restrictive.[3] In a world lacking both computational power and data, and in lieu of powerful theorems, it is little wonder that workaday macroeconomists decided to work with Solow's parsimonious aggregate production function instead. After all, it was easy to work with and only needed a sparing amount of data to be calibrated, typically having just one or two free parameters (the labor share and the elasticity of substitution between capital and labor).[4]

Of course, today's world is awash in an ocean of micro-data and access to computational power is cheap and plentiful, so old excuses no longer apply. Macroeconomic theory must evolve to take advantage of and make sense of detailed micro-level data. This paper is a contribution to this project.

We fully take on board the lessons of the Cambridge-Cambridge controversy and allow for as many factors as necessary to ensure the existence of aggregate production func-

[3]For a good review, see Felipe and Fisher (2003)

[4]The popular specification $Y = AF(A_K K, A_L L)$ with F a CES function described in details in footnote 1 can be entirely calibrated using the labor share ω_L and the elasticity of substitution between capital and labor σ, or even with only the labor share ω_L under the common Cobb-Douglas restriction.

tions.[5] Instead of desperately seeking to aggregate factors, we focus on aggregating over heterogeneous producers in competitive general equilibrium. Under the assumptions of homothetic final demand and no distortions, such aggregation endogenously gives rise to aggregate production functions.[6] The key difference between our approach and that of most of the rest of the literature that follows the Solow-Swan paradigm is that we treat aggregate production functions as endogenous reduced-form objects rather than structural ones. In other words, we do not impose an arbitrary parametric structure on aggregate production functions at the outset and instead derive their properties as a function of deeper structural microeconomic primitives.

Our contribution is to fully characterize these endogenous aggregate production functions, up to the second order, for a general class of competitive disaggregated economies with an arbitrary number of factors and producers, arbitrary patterns of input-output linkages, arbitrary microeconomic elasticities of substitution, and arbitrary microeconomic technology shifters. Our sufficient-statistic formulas lead to general aggregation results expressing the macroeconomic elasticities of substitution between factors and the macroeconomic bias of technical change in terms of microeconomic elasticities of substitution and characteristics of the production network.

The benefits of microfoundations do not require lengthy elaboration. First, they address the Lucas critique by grounding aggregate production functions in deep structural parameters which can be taken to be constant across counterfactuals driven by shocks or policy. Second, they allow us to understand the macroeconomic implications of microeconomic phenomena. Third, they allow to unpack the microeconomic implications of macroeconomic phenomena.

This development can be put in a broader perspective by drawing an analogy with the shifting attitudes of economists towards aggregate consumption functions. In the wake of the Rational Expectations Revolution and the Lucas critique, economists abandoned aggregate consumption functions— functions that postulated a parametric relationship between aggregate consumption and aggregate income without deriving this relationship from microeconomic theory. This has become all the more true with the rise of

[5]Since we do not place any restrictions on the number of factors the economy has, we can recreate the famous counterexamples from the Cambridge capital controversy in our environment. In other words, despite having an aggregate production function, our framework can accommodate the classic Cambridge UK critiques. We show exactly how in Section 5.

[6]We explain later how to generalize our results regarding macroeconomic elasticities of substitution and the macroeconomic bias of technical change to environments with non-homothetic final demand and with distortions.

heterogeneous-agent models following the early contributions of Bewley (1986), Aiyagari (1994), Huggett (1993), and Krusell and Smith (1998). However, the aggregate production function, which does much the same thing on the production side of the economy was left largely unexamined. By deriving an aggregate production function from first-principles, this paper provides microeconomic foundations for the aggregate production function building explicitly on optimizing microeconomic behavior.

We restrict attention to situations where the aggregate production function, a function mapping endowments and technologies to aggregate output, exists, because final demand is homothetic and there are no distortions. The aggregate production function then depends on the structure of final demand. It is possible to define an alternative notion, the aggregate distance function, which separates technology from final demand. The aggregate distance function must then be combined with final demand to compute general-equilibrium comparative statics. Our object of interest is the aggregate production function, which takes these two steps at once. This choice, which is guided by our focus on general-equilibrium comparative statics, is without loss of generality. Indeed, it is possible to use our results for the aggregate production function to characterize the aggregate distance function by using different specifications of final demand.

Aggregate production functions may fail to exist if there is no single quantity index corresponding to final output; this happens if final demand is non-homothetic either because there is a representative agent with non-homothetic preferences or because there are heterogeneous agents with different preferences. Furthermore, aggregate production functions also fail to exist in economies with distortions. Extended notions of aggregate production functions with distortions and non-homothetic final demand can be defined. However, they are less useful in the sense that their properties cannot anymore be tied to interesting observables: their first and second derivatives do not correspond to factor shares, elasticities of substitution between factors, and bias of technical change.

In this paper, we confine ourselves to economies with homothetic final demand and without distortions. In other papers (see Baqaee and Farhi, 2019c, 2018), we have developed an alternative "propagation-equations" methodology to cover economies with non-homothetic final demand and with distortions. These propagation equations generalize equations (5) and (6) in Proposition 2 and equations (8) and (9) in Proposition 7. They fully characterize the elasticities of sales shares and factor shares to factor supplies, factor prices, and technology shocks. They can be used along the exact same lines as in this paper to express the macroeconomic elasticities of substitution between factors and the macroeconomic bias of technical change as a function of microeconomic primitives. This shows precisely how to extend our results to economies with non-homothetic final demand and with distortions.

The outline of the paper is as follows. In Section 2 we set up the basic model, introduce the aggregate cost and production functions as dual ways of representing an economy's production possibilities, and define the notions of macroeconomic elasticities of substitution between factors and of the bias of technical change. In Section 3, we define and characterize the properties of aggregate cost functions for the case of nested-CES economies. In Section 4, we define and characterize the properties of aggregate production functions for the case of nested-CES economies. In Section 5, we review some classic aggregation theorems and provide new ones. We revisit the Cambridge-Cambridge controversy, and represent some of the classic arguments via our framework and language. In Section 6, we provide some simple theoretical examples to illustrate the results developed in Sections 3 and 4: Hicksian and non-Hicksian examples with two and three factors; an example showing how to capture Houthakker (1955) within our framework; and an example of factor-biased technical change in a task-based model. We also present a simple quantitative application to capital-skill complementarity à la Griliches (1969) in the US economy, taking into account the multiplicity of sectors and their input-output linkages. We put it to use to revisit the analysis in Krusell et al. (2000) of the role of these complementarities in the evolution over time of the skill premium. In Section 7, we review several extensions. First, we generalize the results of Sections 3 and 4 to non-nested-CES economies with two simple tricks. Second, we explain how to use our results to separate technology from final demand by characterizing the aggregate distance and associated cost functions. Third, we explain how to generalize our results to economies where final demand is non-homothetic and with distortions. We conclude in Section 8.

2 Setup

In this section, we setup the model and notation, define the equilibrium, the aggregate production, and the aggregate cost function.

2.1 Environment

The model has a set of producers N, and a set of factors F with supply functions L_f. We write $N + F$ for the union of these two sets. With some abuse of notation, we also denote by N and F the cardinalities of these sets. What distinguishes goods from factors is the fact that goods are produced from factors and goods, whereas factors are produced ex nihilo. The output of each producer is produced using intermediate inputs and factors, and is sold as an intermediate good to other producers and as a final good.

Final demand is a constant-returns-to-scale aggregator

$$Y = \mathcal{D}_0(c_1, \ldots, c_N), \tag{1}$$

where c_i represents the use of good i in final demand and Y is real output.

Each good i is produced with some constant-returns- to-scale production function. Hence, we can write the production function of each producer as

$$y_i = A_i F_i(x_{i1}, \cdots, x_{iN}, L_{i1}, \cdots, L_{iF}) \tag{2}$$

where y_i is the total output of i, x_{ij} is the use of input j, and L_{if} is the use of factor f. The variable A_i is a Hicks-neutral productivity shifter. We will sometimes use the unit-cost function $A_i^{-1} C_i(p_1, \cdots, p_N, w_1, \cdots, w_F)$ associated with the production function F_i.

Finally, the economy-wide resource constraints for goods j and factors f are given by:

$$c_j + \sum_{i \in N} x_{ij} = y_j, \tag{3}$$

$$\sum_{i \in N} L_{if} = L_f. \tag{4}$$

This framework is more general than it might appear. First, although we have assumed constant-returns-to-scale production functions, our analysis also covers the case of decreasing-returns-to-scale production functions by adding producer-specific fixed factors.[7,8,9] Similarly, although we have assumed that, at the level of individual producers, technical change is Hicks neutral, our analysis also covers the case of biased factor- or input-augmenting technical change: for example, to capture factor-f-augmenting technical change for firm i, we simply introduce a new fictitious producer which linearly transforms factor f into factor f for firm i and study a Hicks-neutral technology shock to this fictitious producer.[10] Finally, although we refer to each producer as producing one good, our framework actually allows for joint production by multi-product producers: for example, to capture a producer i producing goods i and i' using intermediate inputs and factors, we represent good i' as an input entering negatively in the production and cost functions for good i.[11]

2.2 Feasible and Competitive Equilibrium Allocations

We first define feasible allocations.

Definition. (Feasible Allocations) A feasible allocation is a set of intermediate input choices x_{ij}, factor input choices L_{if}, outputs y_i, final demands c_i, and real output Y, such that (1), (2), (3), and (4) hold.

Next we define equilibrium allocations. Equilibrium allocations are feasible allocations which arise as part of a competitive equilibrium.

Definition. (Equilibrium Allocations) An equilibrium allocation is a set of prices p_i and w_f for goods and factors, intermediate input choices x_{ij}, factor input choices L_{if}, outputs

[7] This was an observation made by McKenzie (1959). See Section 6.3 for a concrete example in the model of Houthakker (1955).

[8] Note that this flexibility also allows us to capture different renditions of assignment/sorting models (see Sattinger, 1993, for a survey). In these models: there are distributions of workers and tasks of different types; output in a given task depends on both the type of the worker and the type of the task; the output of different tasks can be complements or substitutes; and tasks and workers may operate under decreasing returns to scale, limiting, for example, the number of workers per task. These models can be captured within our framework by treating different types of workers as different factors, different tasks as different producers, and allowing for producer-specific fixed factor to capture decreasing returns at the task level.

[9] Our formulas can also in principle be applied, with increasing-returns to scale, but only to some limited extent, by allowing these quasi-fixed factors to be local "bads" receiving negative payments over some range, but care must be taken because it introduces non-convexities in the cost-minimization over variable inputs, and our formulas only apply when variable input demand changes smoothly.

[10] We generalize our results to factor-biased but not necessarily factor-augmenting technical change at the producer level in Section 7.

[11] To satisfactorily capture such features, one probably needs to go beyond the nested-CES case on which we focus for a large part of the paper, and use instead the non-parametric generalization to arbitrary economies provided in Section 7.

y_i, final demands c_i, and real output Y, such that: final demand maximizes Y subject to (1) and to the budget constraint $\sum_{i=1}^{N} p_i c_i = \sum_{f=1}^{F} w_f L_f$; each producer i maximizes its profits $p_i y_i - \sum_{j \in N} p_j x_{ij} - \sum_{f \in F} w_f L_{if}$ subject to (2), taking prices p_j and wages w_f as given; the markets for all goods i and factors f clear so that (3) and (4) hold. Instead of fixing factor supplies L_f, we can also define feasible and equilibrium allocations for given factor prices w_f and level of income E allocated to final demand.

The welfare theorems apply in our environment. Equilibrium allocations are efficient and coincide with the solutions of the planning problems introduced below, which define the aggregate production and cost functions. We will make use of these theorems to go back and forth between those properties most easily seen using the equilibrium decentralization and those that arise more naturally using the planning problem.

Going forward, and to make the exposition more intuitive, we slightly abuse notation in the following way. For each factor f, we interchangeably use the notation w_f or p_{N+f} to denote its wage, the notation L_{if} or $x_{i(N+f)}$ to denote its use by producer i, and the notation L_f or y_f or to denote total factor supply. We define final demand as an additional good produced by producer 0 according to the final demand aggregator. We interchangeably use the notation c_{0i} or x_{0i} to denote the consumption of good i in final demand. We write $1 + N$ for the union of the sets $\{0\}$ and N, and $1 + N + F$ for the union of the sets $\{0\}$, N, and F.

2.3 Aggregate Production and Cost Functions

The aggregate production function is defined as the solution of the following planning problem:

$$F(L_1, \cdots, L_F, A_1, \cdots, A_N) = \max Y$$

subject to (1), (2), (3), and (4). It is homogeneous of degree one in the factor supplies $L_1, \cdots, L_F$. As already discussed above, this production function also indexes the equilibrium level of real output as a function of productivity shocks A_i and factor supplies L_f.

The aggregate cost function is defined as the solution of the dual planning problem which seeks to minimize the expenditure necessary to achieve real output Y given factor prices w_f:

$$C(w_1, \cdots, w_F, A_1, \cdots A_N, Y) = \min E$$

subject to (1), (2), (3), and $E = \sum_{f \in F} w_f L_f$. It is homogeneous of degree one in the factor prices $w_1, \cdots, w_F$. The aggregate cost function is also homogeneous of degree one in aggregate output Y so that we can write it as $YC(w_1, \ldots, w_F, A_1, \ldots, A_N)$, where with some abuse of notation, C now denotes the aggregate unit-cost function. Most of the results in the rest of the paper characterize the log derivatives of the aggregate cost function with respect to productivities or factor prices, which coincide with the corresponding log derivatives of the aggregate unit-cost function, and so both can be used interchangeably. To fix ideas, the reader can focus on the aggregate cost function.

The primary difference between the aggregate production function and the aggregate cost function is that the former takes the factor quantities as given, while the latter takes the factor prices as given.

The goal of this paper is to characterize the aggregate production and cost functions up to the second order as a function of microeconomic primitives such as microeconomic elasticities of substitution and the input-output network. Propositions 1 and 6 characterize the Jacobians (first derivatives) and Propositions 2 and 7 the Hessians (second derivatives) of the aggregate production and cost functions.

In economic terms, this means that we seek to characterize not only macroeconomic marginal products of factors and factor demands (first-order properties) but also macroeconomic elasticities of substitution between factors and the sensitivities of marginal products of factors and factor demands to technical change (second-order properties).

As mentioned before, we have put ourselves under conditions where the existence of an "aggregate output" good can be taken for granted because final demand is homothetic. Given our definitions, the properties of the aggregate production and cost functions depend on final demand. One might want to consider alternative notions which remove this dependence: the aggregate distance and associated cost functions. One might also want to study economies where there is no aggregate output good because final demand is not homothetic, or economies with distortions. Our results can be used to fully and precisely accommodate these desiderata. Detailed explanations can be found in Sections A.3 and A.4.

2.4 Macroeconomic Elasticities of Substitution Between Factors

As is well known, there is no unambiguous way to generalize the standard Hicksian notion of elasticity of substitution between factors (Hicks, 1932) when there are more than two factors, and several concepts have been proposed in the literature.

Invariably, all competing definitions of the elasticity of substitution are computed via the Jacobian and Hessian of a function. Since we characterize both of these in general, our results can be used to compute all the different notions of the elasticity of substitution. In this paper, we follow Blackorby and Russell (1989) who advocate using the definition due to Morishima (1967). They argue that Morishima Elasticities of Substitution (MESs) are appealing because they extend the standard Hicksian notion while preserving some of its desirable properties: an MES is a measure of the inverse-curvature of isoquants; it is a sufficient statistic for the effect on relative factor shares of changes in relative factor prices; it is a log derivative of a quantity ratio to a price ratio.[12]

Definition. (MESs for the Aggregate Production Function) The MES σ^F_{fg} between factors f and g in the aggregate production function is defined as

$$\frac{1}{\sigma^F_{fg}} = \frac{d\log(\frac{dF}{dL_f}/\frac{dF}{dL_g})}{d\log(L_g/L_f)} = 1 + \frac{d\log(\frac{d\log F}{d\log L_f}/\frac{d\log F}{d\log L_g})}{d\log(L_g/L_f)}.$$

Definition. (MESs for the Aggregate Cost Function) The MES σ^C_{fg} between factors f and g in the aggregate cost function is defined as

$$\sigma^C_{fg} = \frac{d\log(\frac{dC}{dw_f}/\frac{dC}{dw_g})}{d\log(w_g/w_f)} = 1 + \frac{d\log(\frac{d\log C}{d\log w_f}/\frac{d\log C}{d\log w_g})}{d\log(w_g/w_f)}.$$

Note that the ratios $(dF/dŁ_f)/(dF/dL_g)$ and $(d\log F/d\log L_f)/(d\log F/d\log L_g)$ are homogeneous of degree zero in $L_1, \cdots, L_F$. Similarly, the ratios $(dC/dw_f)/(dC/dw_g)$

[12]Stern (2010) points out that while the MES in cost do characterize the inverse-curvature of the corresponding constant-output isoquants, those in production do not. In the production function case, he defines the symmetric elasticity of complementarity to be the inverse-curvature of the constant-output isoquants, and shows that its inverse, which measures the curvature of the constant-output isoquants, is symmetric and can easily be recovered as a share-weighted harmonic average $(\Lambda_f + \Lambda_g)/(\Lambda_f/\sigma^F_{fg} + \Lambda_g/\sigma^F_{gf})$ of the MESs in production. This concept is the dual of the shadow elasticity of substitution put forth by McFadden (1963), which is symmetric and which can be recovered as share-weighted arithmetic average $(\Lambda_f\sigma^C_{fg} + \Lambda_g\sigma^C_{gf})/(\Lambda_f + \Lambda_g)$ of the MESs in costs. We will focus on characterizing MESs for the aggregate production and cost functions.

and $(d\log C/d\log w_f)/(d\log C/d\log w_g)$ are homogeneous of degree zero in $w_1,\cdots,w_F$. These definitions exploit this homogeneity to write these ratios as functions of $L_1/L_f,\cdots,L_F/L_f$ and $w_1/w_f,\cdots,w_F/w_f$, respectively. Therefore, underlying the definition of σ^F_{fg} are variations in L_g/L_f, holding L_h/L_f constant for $h\neq g$, i.e. variations in L_g, holding L_h constant for $h\neq g$. Similarly, underlying the definition of σ^C_{fg} are variations in w_g/w_f, holding w_h/w_f constant for $h\neq g$, i.e. variations in w_g, holding w_h constant for $h\neq g$.

As we shall see below in Propositions 1 and 6, $d\log F/d\log L_h$ and $d\log C/d\log w_h$ are equal to the factor shares Λ_h in the competitive equilibria of the corresponding economies. MESs therefore pin down the elasticities of relative factor shares to relative factor supplies or relative factor prices:

$$1-\frac{1}{\sigma^F_{fg}}=-\frac{d\log(\Lambda_f/\Lambda_g)}{d\log(L_g/L_f)}\quad\text{and}\quad\sigma^C_{fg}-1=\frac{d\log(\Lambda_f/\Lambda_g)}{d\log(w_g/w_f)}.$$

Similarly, dF/dL_h is equal to the wage rate w_h and dC/dw_h to the factor demand per unit of output L_h in the competitive equilibria of the corresponding economies, which can be viewed as homogeneous-of-degree-zero functions of $L_1,\cdots,L_F$ and $w_1,\cdots,w_F$ respectively. MESs therefore pin down the elasticities of factor prices to factor supplies and of factor demands to factor prices:

$$\frac{1}{\sigma^F_{fg}}=\frac{d\log(w_f/w_g)}{d\log(L_g/L_f)}\quad\text{and}\quad\sigma^C_{fg}=\frac{d\log(L_f/L_g)}{d\log(w_g/w_f)}.$$

MESs between factors in the aggregate production and cost functions can be directly expressed as a function of the Jacobians and Hessians of these functions:

$$1-\frac{1}{\sigma^F_{fg}}=\frac{d^2\log F/(d\log L_g)^2}{d\log F/d\log L_g}-\frac{d^2\log F/(d\log L_f d\log L_g)}{d\log F/d\log L_f},$$
$$\sigma^C_{fg}-1=\frac{d^2\log C/(d\log w_f d\log w_g)}{d\log C/d\log w_f}-\frac{d^2\log C/(d\log w_g)^2}{d\log C/d\log w_g}.$$

MESs between factors in the aggregate production function are typically not symmetric so that $\sigma^F_{fg}\neq\sigma^F_{gf}$ and $\sigma^C_{fg}\neq\sigma^C_{gf}$ in general. Moreover, MESs between factors in the aggregate production and cost functions are typically not equal to each other, so $\sigma^F_{fg}\neq\sigma^C_{gf}$ in general.

The "Hicksian" case where there are only two factors of production f and g is special in this regard since in this case, the MESs for the cost and production function are the same, and symmetric, so that we get $\sigma^F_{fg} = \sigma^F_{gf}$, $\sigma^C_{fg} = \sigma^C_{gf}$, and $\sigma^F_{fg} = \sigma^C_{fg}$. The proof is standard and can be found in Hicks (1932) and in Russell (2017) for example.

Consider the case where the aggregate production function and the associated aggregate cost function are of the CES form with

$$F(L_1, \ldots, L_N, A) = \bar{Y}\frac{A}{\bar{A}}\left(\sum_{i=1}^{N}\omega_i\left(\frac{L_i}{\bar{L}_i}\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}},$$

$$C(w_1, \ldots, w_N, A) = \frac{\bar{A}}{A}\left(\sum_{i=1}^{N}\omega_i\left(\frac{w_i}{\bar{w}_i}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}},$$

where bar variables correspond to some particular point and ω_i denotes the share of factor i at this point. Then with our definitions, the MESs in the aggregate production and cost functions between factor f and factor g are given by $\sigma^F_{fg} = \sigma^C_{fg} = \sigma$.

More generally, if the aggregate production and cost functions are of the nested-CES form, and if two factors belong to the same CES nest, then the MES between these two factors is equal to the elasticity of substitution of the nest; more generally, if two factors enter together with other factors only through a nested-CES sub-aggregate, then the MES between these two factors is only a function of the elasticities of substitution in the nested-CES sub-aggregate.

However, even when the economy with disaggregated production is of the nested-CES form as described in Section 2.7, the aggregate production and cost functions that describe its production possibility frontier are typically not of the nested-CES form except in simple cases with limited heterogeneity and simple input-output network structures. MESs between factors in the aggregate production and cost functions are macroeconomic elasticities of substitution. They incorporate general equilibrium effects and typically do not coincide with any microeconomic elasticity of substitution.

Our results in Propositions 3 and 8 below deliver formulas for the MESs between factors as a function of microeconomic primitives such as microeconomic elasticities of substitution and the input-output network.

2.5 Macroeconomic Bias of Technical Change

We now present our definitions of the macroeconomic bias of technical change. These definitions generalize the definitions proposed by (Hicks, 1932) to the case of multiple factors. We present these definitions directly in terms of the Jacobians and Hessians of the aggregate production and cost functions. We later relate them to the elasticities of relative factor shares to technology shocks.

Definition. (Bias of Technical Change for the Aggregate Production Function) The bias B^F_{fgj} in the aggregate production function towards factor f vs. factor g of technical change driven by a technology shock to producer j is defined as

$$\frac{B^F_{fgj}}{1+B^F_{fgj}} = \frac{d\log(\frac{d\log F}{d\log L_f}/\frac{d\log F}{d\log L_g})}{d\log A_j}.$$

Definition. (Bias of Technical Change for the Aggregate Cost Function) The bias B^C_{fgj} in the aggregate cost function towards factor f vs. factor g of technical change driven by a technology shock to producer j is defined as

$$B^C_{fgj} = \frac{d\log(\frac{d\log C}{d\log w_f}/\frac{d\log C}{d\log w_g})}{d\log A_j}.$$

As already alluded to, and as we shall see below in Propositions 1 and 6, $d\log F/d\log L_h$ and $d\log C/d\log w_h$ are equal to the factor shares Λ_h in the competitive equilibria of the corresponding economies. The macroeconomic biases of technical change in the aggregate production and cost functions therefore pin down the elasticities with respect to technology shocks of relative factor shares as well as of relative factor prices and of relative factor demands, holding respectively factor supplies or factor prices constant :[13]

[13]We can use these measures to compute a measure of bias of technical change towards one factor instead of towards one factor vs. another by defining

$$B^F_{fj} = \sum_{g\in F}\frac{d\log F}{d\log L_g}B^F_{fgj} = \sum_{g\in F}\Lambda_g B^F_{fgj} = \frac{d\log\Lambda_f}{d\log A_j} = \frac{d\log w_f}{d\log A_j} - \lambda_j$$

and

$$B^C_{fj} = \sum_{g\in F}\frac{d\log C}{d\log w_g}B^C_{fgj} = \sum_{g\in F}\Lambda_g B^C_{fgj} = \frac{d\log\Lambda_f}{d\log A_j} = \frac{d\log L_f}{d\log A_j} - \lambda_j,$$

holding respectively factor supplies or factor prices constant.

$$\frac{B^F_{fgj}}{1+B^F_{fgj}} = \frac{d\log(\Lambda_f/\Lambda_g)}{d\log A_j} = \frac{d\log(w_f/w_g)}{d\log A_j} \quad \text{and} \quad B^C_{fgj} = \frac{d\log(\Lambda_f/\Lambda_g)}{d\log A_j} = \frac{d\log(L_f/L_g)}{d\log A_j}.$$

Technological bias in the aggregate production and cost functions can be directly expressed as a function of the Jacobians and Hessians of these functions:

$$\frac{B^F_{fgj}}{1+B^F_{fgj}} = \frac{d^2\log F/(d\log A_j d\log L_f)}{d\log F/d\log L_f} - \frac{d^2\log F/(d\log A_j d\log L_g)}{d\log F/d\log L_g},$$
$$B^C_{fgj} = \frac{d^2\log C/(d\log A_j d\log w_f)}{d\log C/d\log w_f} - \frac{d^2\log C/(d\log A_j d\log w_g)}{d\log C/d\log w_g}.$$

Even in the Hicksian case where there are only two factors, the biases of technology in the aggregate production and cost functions do not necessarily coincide so that $B^F_{fgj} \neq B^C_{fgj}$ in general.

Consider the case where the aggregate production and cost functions are of the CES form with

$$F(L_1,\ldots,L_N,A_1,\ldots,A_N) = \bar{Y}\left(\sum_{i=1}^N \omega_i \left(\frac{A_i}{\bar{A}_i}\frac{L_i}{\bar{L}_i}\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}},$$

$$C(w_1,\ldots,w_N,A_1,\ldots,A_N) = \left(\sum_{i=1}^N \omega_i \left(\frac{\bar{A}_i}{A_i}\frac{w_i}{\bar{w}_i}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}},$$

where bar variables correspond to some particular point, ω_i denotes the share of factor i at this point, and A_i is a technological shock augmenting factor i. Then with our definitions, the biases in the aggregate production and cost functions towards factor f vs. factor g of technical change driven by a technology shock augmenting factor j are given $B^F_{fgj} = B^C_{fgj} = \sigma - 1$ if $j = f$, $B^F_{fgj} = B^C_{fgj} = -(\sigma-1)$ if $j = g$, and $B^F_{fgj} = B^C_{fgj} = 0$ otherwise.

More generally, if the aggregate production and cost functions are of the nested-CES form with factor-augmenting technical change, and if two factors belong to the same CES nest, then the bias towards the first factor vs. the second factor of a technology shock is equal to (minus) the elasticity of substitution minus one of the nest if the technology shock augments the first (second) of the two factors, and zero otherwise; more generally, if two factors enter together with other factors only through a nested-CES sub-aggregate, then the bias between these two factors is nonzero only for technology shocks that augment the factors in the nested-CES sub-aggregate, and then it is only a function of the elasticities of substitution in the nested-CES sub-aggregate.

However, even when the economy with disaggregated production is of the nested-CES form as described in Section 2.7, the aggregate production and cost functions that describe its production possibility frontier are typically not of the nested-CES form with factor-augmenting technical change except in simple cases with limited heterogeneity and simple input-output network structures. The bias of technical change in the aggregate production and cost functions are macroeconomic in nature. They incorporate general equilibrium effects and typically do not coincide with any microeconomic elasticity of substitution minus one.

Our results in Propositions 5 and 10 below deliver formulas for the bias of technical change as a function of microeconomic primitives such as microeconomic elasticities of substitution and the input-output network.

2.6 Input-Output Definitions

To state our results, we require some input-output notation and definitions. We define input-output objects such as input-output matrices, Leontief inverse matrices, and Domar weights. These definitions arise most naturally in the equilibrium decentralization of the corresponding the planning solution (for the aggregate production function or the aggregate cost function respectively).

Input-Output Matrix

We define the input-output matrix to be the $(1 + N + F) \times (1 + N + F)$ matrix Ω whose ijth element is equal to i's expenditures on inputs from j as a share of its total revenues

$$\Omega_{ij} \equiv \frac{p_j x_{ij}}{p_i y_i},$$

Note that input-output matrix Ω includes expenditures by producers on factor inputs as well as expenditures by consumers for final consumption. By Shephard's lemma, Ω_{ij} is also the elasticity of the cost of i to the price of j, holding the prices of all other producers constant.

Leontief Inverse Matrix

We define the Leontief inverse matrix as

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

While the input-output matrix Ω records the *direct* exposures of one producer to another, the Leontief inverse matrix Ψ records instead the *direct and indirect* exposures through the production network. This can be seen most clearly by noting that $(\Omega^n)_{ij}$ measures the weighted sums of all paths of length n from producer i to producer j. By Shephard's lemma, Ψ_{ij} is also the elasticity of the cost of i to the price of j holding fixed the prices of factors but taking into account how the price of all other goods in the economy will change.

GDP and Domar Weights

GDP or nominal output is the total sum of all final expenditures

$$GDP = \sum_{i \in N} p_i c_i = \sum_{i \in N} p_i x_{0i}.$$

We define the Domar weight λ_i of producer i to be its sales share as a fraction of GDP

$$\lambda_i \equiv \frac{p_i y_i}{GDP}.$$

Note that $\sum_{i \in N} \lambda_i > 1$ in general since some sales are not final sales but intermediate sales.

For expositional convenience, for a factor f, we sometimes use Λ_f instead of λ_f. Note that the Domar weight Λ_f of factor f is simply its total income share.

We can also define the vector b to be final demand expenditures as a share of GDP

$$b_i = \frac{p_i c_i}{GDP} = \frac{p_i x_{0i}}{GDP} = \Omega_{0i}.$$

The accounting identity

$$p_i y_i = p_i c_i + \sum_{j \in N} p_i x_{ji} = \Omega_{0i} GDP + \sum_{j \in N} \Omega_{ji} \lambda_j GDP$$

links the revenue-based Domar weights to the Leontief inverse via

$$\lambda' = b'\Psi = b'I + b'\Omega + b'\Omega^2 + \dots .$$

Another way to see this is that the i-th element of $b'\Omega^n$ measures the weighted sum of all paths of length n from producer i to final demand.

2.7 Nested-CES Economies

We call an economy nested CES if all the production functions of all the producers (including final demand) are of the nested-CES form with Hicks-neutral technical change at the level of each nest. Following Baqaee and Farhi (2019a), any nested-CES economy, with an arbitrary number of producers, factors, CES nests, elasticities, and intermediate input use, can be re-written in what we call *standard form*, which is more convenient to study.

A CES economy in standard form is defined by a tuple (ω, θ, F). The $(1+N+F) \times (1+N+F)$ matrix ω is a matrix of input-output parameters. The $(1+N) \times 1$ vector θ is a vector of microeconomic elasticities of substitution. Each good i is produced with the production function

$$\frac{y_i}{\overline{y}_i} = \frac{A_i}{\bar{A}_i} \left(\sum_{j \in 1+N+F} \omega_{ij} \left(\frac{x_{ij}}{\overline{x}_{ij}} \right)^{\frac{\theta_i - 1}{\theta_i}} \right)^{\frac{\theta_i}{\theta_i - 1}},$$

where x_{ij} are intermediate inputs from j used by i. Throughout the paper, variables with over-lines are normalizing constants equal to the values at some initial allocation. We represent final demand Y as the purchase of good 0 from producer 0 producing the final good. For the most part, we assume that $A_0 = \bar{A}_0$ and abstract away from this parameter.[14]

Through a relabelling, this structure can represent any nested-CES economy with an arbitrary pattern of nests and elasticities. Intuitively, by relabelling each CES aggregator to be a new producer, we can have as many nests as desired.

To facilitate the exposition in the paper, and due to their ubiquity in the literature, we present our baseline results for nested-CES economies in Sections 3 and 4. We then

[14]Changes in A_0 are changes in how each unit of final output affects consumer welfare. This is what Hulten and Nakamura (2017) call "output-saving" technical change.

explain how to generalize them for arbitrary economies in Section A.1.[15]

3 Aggregate Cost Functions

In this section, we provide a general characterization of aggregate cost functions up to the second order for nested-CES economies. We refer the reader to Sections 6 and 6.5 for some simple theoretical and quantitative examples, and to Section A.1 for a generalization to non-nested-CES economies.

3.1 First-Order Characterization

The following proposition characterizes the first derivatives (gradient) of the aggregate cost function.

Proposition 1. *(Gradient) The first derivatives of the aggregate cost function are given by the sales shares of goods and factors*

$$\frac{d \log C}{d \log w_f} = \Lambda_f \quad \textit{and} \quad \frac{d \log C}{d \log A_i} = -\lambda_i.$$

The proposition follows directly from Hulten's theorem (Hulten, 1978). It shows that the elasticity of the aggregate cost function C to the price of factor f is given by the share Λ_f of this factor in GDP. Similarly, the elasticity of the aggregate cost function C to the productivity of producer i is given by the negative of the sales share λ_i of this producer in GDP. The proposition is fully general and applies even when the economy is not of the nested-CES form.

Incidentally, Proposition 1 confirms that the aggregate cost function C is homogeneous of degree one in factor prices, since $\sum_{f \in F} \Lambda_f = 1$. It also confirms that C is homogeneous of degree one in aggregate output Y since $\mathrm{d} \log C / \mathrm{d} \log A_0 = 1$.

[15] We assume throughout that all microeconomic elasticities of substitution are finite. Some economic models assume that some of these elasticities are infinite. This implies that substitution is not smooth at the producer level, and raises a number of technical issues having to do with varying patterns of partial or full specialization. Our approach can be used to shed light on these models by viewing them as limiting cases of perhaps models with large but finite elasticities.

3.2 Second-Order Characterization

The following proposition characterizes the second derivatives (Hessian) of the aggregate cost function.

Proposition 2. *(Hessian) The second derivatives of the aggregate cost function are determined by the elasticities of the sales shares of goods and factors*

$$\frac{d^2 \log C}{d \log w_f d \log w_g} = \frac{d\Lambda_f}{d \log w_g},$$
$$\frac{d^2 \log C}{d \log A_j d \log A_i} = -\frac{d\lambda_i}{d \log A_j},$$
$$\frac{d^2 \log C}{d \log A_j d \log w_f} = \frac{d\Lambda_f}{d \log A_j},$$

where the elasticities of the sales shares are given by

$$d \log \lambda_i = \sum_{k \in 1+N} (\theta_k - 1) \frac{\lambda_k}{\lambda_i} Cov_{\Omega^{(k)}} \left(\sum_{j \in N} \Psi_{(j)} d \log A_j - \sum_{g \in F} \Psi_{(g)} d \log w_g, \Psi_{(i)} \right), \tag{5}$$

and the elasticities of the factor shares are given by

$$d \log \Lambda_f = \sum_{k \in 1+N} (\theta_k - 1) \frac{\lambda_k}{\Lambda_f} Cov_{\Omega^{(k)}} \left(\sum_{j \in N} \Psi_{(j)} d \log A_j - \sum_{g \in F} \Psi_{(g)} d \log w_g, \Psi_{(f)} \right). \tag{6}$$

The shares propagation equations (5) and (6) are taken directly from Baqaee and Farhi (2019a). While Baqaee and Farhi (2019a) focuses on the second-order macroeconomic impact of microeconomic shocks $d^2 \log C / (d \log A_j d \log A_i)$, in this paper, we focus instead on $d^2 \log C / (d \log w_f d \log w_g)$, which as we will show in Section 3.3 below, determines the macroeconomic elasticities of substitution between factors, as well as on $d^2 \log C / (d \log A_j d \log w_f)$, which determines the elasticity of factor shares to technical change i.e. the bias of technical change.

Of course, equation (6) is obtained simply by letting $i = f$ in (5). This proposition shows that these equations, which characterize the elasticities of the shares of goods and factors to productivity shocks and factor prices, completely pin down the second derivatives of the aggregate cost function.

In these equations, we make use of the *input-output covariance operator* introduced by

Baqaee and Farhi (2019a):

$$Cov_{\Omega^{(k)}}(\Psi_{(j)}, \Psi_{(i)}) = \sum_{l \in N+F} \Omega_{kl} \Psi_{lj} \Psi_{li} - \left(\sum_{l \in N+F} \Omega_{kl} \Psi_{lj} \right) \left(\sum_{l \in N+F} \Omega_{kl} \Psi_{li} \right), \tag{7}$$

where $\Omega^{(k)}$ corresponds to the kth row of Ω, $\Psi_{(j)}$ to jth column of Ψ, and $\Psi_{(i)}$ to the ith column of Ψ. In words, this is the covariance between the jth column of Ψ and the ith column of Ψ using the kth row of Ω as the distribution. Since the rows of Ω always sum to one for a reproducible (non-factor) good k, we can formally think of this as a covariance, and for a non-reproducible good, the operator just returns 0.

To gain some intuition, consider for example the elasticity $d \log \Lambda_f / d \log w_g$ of the share Λ_f of factor f to the price of factor g in equation (6). Imagine a shock $d \log w_g < 0$ which reduces the wage of factor g. For fixed relative factor prices, every producer k will substitute across its inputs in response to this shock. Suppose that $\theta_k > 1$, so that producer k substitutes expenditure *towards* those inputs l that are more reliant on factor g, captured by Ψ_{lg}, and the more so, the higher $\theta_k - 1$. Now, if those inputs are also more reliant on factor f, captured by a high $Cov_{\Omega^{(k)}}\left(\Psi_{(g)}, \Psi_{(f)}\right)$, then substitution by k will increase expenditure on factor f and hence the income share of factor f. These substitutions, which happen at the level of each producer k, must be summed across producers. The intuition for $d \log \Lambda_f / d \log A_j$ in equation (6) as well as for $d \log \lambda_i / d \log w_g$ and $d \log \lambda_i / d \log A_j$ in equation 5 is similar.

3.3 Macroeconomic Elasticities of Substitution Between Factors

We can leverage Proposition 2 to characterize the macroeconomic elasticities of substitution between factors in the aggregate cost function.

Proposition 3. *(MESs) The MESs between factors in the aggregate cost function are given by*

$$\sigma^C_{fg} = \sum_{k \in 1+N} \theta_k \lambda_k Cov_{\Omega^{(k)}}(\Psi_{(g)}, \Psi_{(g)}/\Lambda_g - \Psi_{(f)}/\Lambda_f),$$

where

$$\sum_{k \in 1+N} \lambda_k Cov_{\Omega^{(k)}}(\Psi_{(g)}, \Psi_{(g)}/\Lambda_g - \Psi_{(f)}/\Lambda_f) = 1.$$

This proposition shows that MESs σ^C_{fg} between factors in the aggregate cost function are *weighted averages* of the microeconomic elasticities of substitution θ_k in production with weights given by sufficient statistics of the input-output network $\lambda_k Cov_{\Omega^{(k)}}(\Psi_{(g)}, \Psi_{(g)}/\Lambda_g - \Psi_{(f)}/\Lambda_f)$. These weights capture the change in expenditure for factor f vs. g from substitution by producer k in response to a change in the price of factor f.

This implies the following network-irrelevance result, already uncovered in Baqaee and Farhi (2019a), in the knife-edge case where all the microeconomic elasticities of substitution are identical.

Proposition 4. *(Network Irrelevance) If all microeconomic elasticities of substitution θ_k are equal to the same value $\theta_k = \theta$, then MESs σ^C_{fg} between factors in the aggregate cost function are also equal to that value $\sigma^C_{fg} = \theta$.*

3.4 Macroeconomic Bias of Technical Change

We can also leverage Proposition 2 to characterize the macroeconomic bias of technical change in the aggregate cost function.

Proposition 5. *(Bias of Technical Change) The biases towards one factor vs. another of the different technology shocks in the aggregate cost function are given by*

$$B^C_{fgj} = \sum_{k \in 1+N} (\theta_k - 1) \lambda_k Cov_{\Omega^{(k)}}(\Psi_{(j)}, \Psi_{(f)}/\Lambda_f - \Psi_{(g)}/\Lambda_g).$$

This proposition shows that biases B^C_{fgj} are weighted sums of the departures from one $\theta_k - 1$ of the microeconomic elasticities of substitution with weights given by sufficient statistics of the input-output network $\lambda_k Cov_{\Omega^{(k)}}(\Psi_{(j)}, \Psi_{(f)}/\Lambda_f - \Psi_{(g)}/\Lambda_g)$. These weights capture the change in expenditure for factor f vs. g as a result substitution by producer k in response to a technology shock to producer j.

The network-irrelevance result for MESs in the aggregate cost function stated in Proposition 4 does not extend to the bias of technical change. In general, the network matters for the bias of technical change, even when all the microeconomic elasticities of substitution θ_k are identical. The Cobb-Douglas is the one case where it doesn't: when all the microeconomic elasticities θ_k are unitary so that $\theta_k = 1$, technical change in unbiased with $B^C_{fgj} = 0$ for all f, g, and j, no matter what the structure of the network is.

4 Aggregate Production Functions

In this section, we provide a general characterization of aggregate production functions up to the second order for nested-CES economies. We refer the reader to Sections 6 and 6.5 for some simple theoretical and quantitative examples, and to Section A.1 for a generalization to non-nested-CES economies.

4.1 First-Order Characterization

The following proposition characterizes the first derivatives (gradient) of the aggregate production function.

Proposition 6. *(Gradient) The first derivatives of the aggregate production function are given by the sales shares of goods and factors*

$$\frac{d \log F}{d \log L_f} = \Lambda_f \quad \textit{and} \quad \frac{d \log F}{d \log A_i} = \lambda_i.$$

The proposition follows directly from Hulten's theorem (Hulten, 1978). It shows that the elasticity of the aggregate production function F to the supply of factor f is given by the share Λ_f of this factor in GDP. Similarly, the elasticity of the aggregate production function F to the productivity of producer i is given by the sales share λ_i of this producer in GDP. The proposition is fully general and applies even when the economy is not of the nested-CES form.

Once again, Proposition 6 confirms that the aggregate production function is homogeneous of degree one with respect to factor quantities since $\sum_{f \in F} \Lambda_f = 1$.

4.2 Second-Order Characterization

The following proposition characterizes the second derivatives (Hessian) of the aggregate production function.

Proposition 7. *(Hessian) The second derivatives of the aggregate production function are given* *by the elasticities of the sales shares of goods and factors*

$$\frac{d^2 \log F}{d \log L_f d \log L_g} = \frac{d\Lambda_f}{d \log L_g},$$
$$\frac{d^2 \log F}{d \log A_j d \log A_i} = \frac{d\lambda_i}{d \log A_j},$$
$$\frac{d^2 \log F}{d \log A_j d \log L_f} = \frac{d\Lambda_f}{d \log A_j},$$

where the elasticities of the sales shares are given by

$$d \log \lambda_i = \sum_{k \in 1+N} (\theta_k - 1) \frac{\lambda_k}{\lambda_i} Cov_{\Omega^{(k)}} (\sum_{j \in N} \Psi_{(j)} d \log A_j + \sum_{g \in F} \Psi_{(g)} d \log L_g, \Psi_{(i)})$$
$$- \sum_{h \in F} \sum_{k \in 1+N} (\theta_k - 1) \frac{\lambda_k}{\lambda_i} Cov_{\Omega^{(k)}} (\Psi_{(h)}, \Psi_{(i)}) d \log \Lambda_h, \quad (8)$$

and where the elasticities of the factor shares solve the following system of linear equations

$$d \log \Lambda_f = \sum_{k \in 1+N} (\theta_k - 1) \frac{\lambda_k}{\Lambda_f} Cov_{\Omega^{(k)}} (\sum_{j \in N} \Psi_{(j)} d \log A_j + \sum_{g \in F} \Psi_{(g)} d \log L_g, \Psi_{(f)})$$
$$- \sum_{h \in F} \sum_{k \in 1+N} (\theta_k - 1) \frac{\lambda_k}{\Lambda_f} Cov_{\Omega^{(k)}} (\Psi_{(h)}, \Psi_{(f)}) d \log \Lambda_h. \quad (9)$$

The shares propagation equations (8) and (9) are taken directly from Baqaee and Farhi (2019a). While Baqaee and Farhi (2019a) focuses on the second-order macroeconomic impact of microeconomic shocks $d^2 \log F / (d \log A_j d \log A_i)$, in this paper, we focus instead on $d^2 \log F / (d \log L_f d \log L_g)$, which as we will show in Section 4.3 below, determines the macroeconomic elasticities of substitution between factors, as well as on $d^2 \log F / (d \log A_j d \log w_f)$, which determines the elasticity of factor shares to technical change i.e. the bias of technical change.

The difference with the characterization of the second-order aggregate cost function in Section 3.2 is that: the elasticities of the factor shares show up in equation (8) for the elasticities of the sales shares; the elasticities of the sales shares are now given by a system of linear equations. As we shall see, this is because shocks trigger changes in relative demand for factors, which given fixed factor supplies, lead to changes in factor prices.

To gain some intuition, consider for example the vector of elasticities $d \log \Lambda / d \log L_g$

of factor shares to the supply of factor g. Note that as observed in Baqaee and Farhi (2019a), we can rewrite the system of linear factor share propagation equations (9) as

$$\frac{d \log \Lambda}{d \log L_g} = \Gamma \frac{d \log \Lambda}{d \log L_g} + \delta_{(g)}, \tag{10}$$

with

$$\Gamma_{fh} = - \sum_{k \in 1+N} (\theta_k - 1) \frac{\lambda_k}{\Lambda_f} Cov_{\Omega^{(k)}} \left(\Psi_{(h)}, \Psi_{(f)} \right),$$

and

$$\delta_{fg} = \sum_{k \in 1+N} (\theta_k - 1) \frac{\lambda_k}{\Lambda_f} Cov_{\Omega^{(k)}} \left(\Psi_{(g)}, \Psi_{(f)} \right).$$

We call δ the *factor share impulse matrix*. Its gth column encodes the direct or first-round effects of a shock to the supply of factor g on factor income shares, taking relative factor prices as given. We call Γ the *factor share propagation matrix*. It encodes the effects of changes in relative factor prices on factor income shares, and it is independent of the source of the shock g.

Consider a shock $d \log L_g > 0$ which increases the supply of factor g. If we fix relative factor shares, the relative price of this factor declines by $-d \log L_g$. Every producer k will substitute across its inputs in response to this shock. Suppose that $\theta_k > 1$, so that producer k substitutes expenditure *towards* those inputs l that are more reliant on factor g, captured by Ψ_{lg}, and the more so, the higher $\theta_k - 1$. Now, if those inputs are also more reliant on factor f, captured by a high $Cov_{\Omega^{(k)}} \left(\Psi_{(g)}, \Psi_{(f)} \right)$, then substitution by k will increase expenditure on factor f and hence the income share of factor f. These substitutions, which happen at the level of each producer k, must be summed across producers.

This first round of changes in the demand for factors triggers changes in relative factor prices which then sets off additional rounds of substitution in the economy that we must account for, and this is the role Γ plays. For a given set of factor prices, the shock to g affects the demand for each factor, hence factor income shares and in turn factor prices, as measured by the $F \times 1$ vector $\delta_{(g)}$ given by the gth column of δ. These changes in factor prices then cause further substitution through the network, leading to additional changes in factor demands and prices. The impact of the change in the relative price of factor h on the share of factor f is measured by the fhth element of the $F \times F$ matrix Γ. The movements in factor shares are the fixed point of this process, i.e. the solution of equation

(10):

$$\frac{d \log \Lambda}{d \log L_g} = (I - \Gamma)^{-1} \delta_{(g)},$$

where I is the $F \times F$ identity matrix.

The intuition for the elasticities of factor share to productivity shocks $d \log \Lambda / d \log A_j$ in equation (9) and for the elasticities of sales shares of goods to factor supplies $d \log \lambda / d \log L_g$ and to productivities $d \log \lambda / d \log A_j$ in equation (8) are similar.

4.3 Macroeconomic Elasticities of Substitution Between Factors

As in Section 3.3, we can leverage Proposition 7 to characterize the macroeconomic elasticities of substitution between factors in the aggregate production function.

Proposition 8. *(MESs) The MESs between factors in the aggregate production function are given by*

$$1 - \frac{1}{\sigma^F_{fg}} = (I_{(g)} - I_{(f)})'(I - \Gamma)^{-1} \delta_{(g)},$$

where Γ *is the* $F \times F$ *factor share propagation matrix defined by*

$$\Gamma_{hh'} = - \sum_{k \in 1+N} (\theta_k - 1) \lambda_k Cov_{\Omega^{(k)}} \left(\Psi_{(h')}, \Psi_{(h)} / \Lambda_h \right),$$

δ *is the* $F \times F$ *factor share impulse matrix defined by*

$$\delta_{hh'} = \sum_{k \in 1+N} (\theta_k - 1) \lambda_k Cov_{\Omega^{(k)}} \left(\Psi_{(h')}, \Psi_{(h)} / \Lambda_h \right)$$

$\delta_{(g)}$ *is its gth column,* I *is the* $F \times F$ *identity matrix, and* $I_{(f)}$ *and* $I_{(g)}$*are its fth and gth columns.*

In Section 3.3, we showed that the MES σ^C_{fg} between factors in the aggregate cost function are *weighted averages* of the microeconomic elasticities of substitution θ_k in production with weights given by sufficient statistics $\lambda_k Cov_{\Omega^{(k)}}(\Psi_{(g)}, \Psi_{(g)} / \Lambda_g - \Psi_{(f)} / \Lambda_f)$ of the input-output network. For the aggregate production function, the MESs σ^F_{fg} between factors are still determined by microeconomic elasticities of substitution θ_k and by sufficient statistics $\lambda_k Cov_{\Omega^{(k)}}(\Psi_{(h')}, \Psi_{(h)} / \Lambda_h)$ of the input-output network. However, they are no longer weighted averages of the microeconomic elasticities of substitution, and they depend on a longer list of input-output network sufficient statistics. In fact, σ^F_{fg} is now a *nonlinear* function of the sufficient statistics $(\theta_k - 1) \lambda_k Cov_{\Omega^{(k)}}(\Psi_{(h')}, \Psi_{(h)} / \Lambda_h)$.

There are two special cases where σ^F_{fg} becomes a weighted average of the microeconomic elasticities θ_i, The first case is the "Hicksian" case when there are only two factors. The second case is when all the microeconomic elasticities of substitution are identical, which follows from the following network-irrelevance result established in uncovered in Baqaee and Farhi (2019a).

Proposition 9. *(Network Irrelevance) If all microeconomic elasticities of substitution θ_k are equal to the same value $\theta_k = \theta$, then MESs σ^F_{fg} between factors in the aggregate production function are also equal to that value $\sigma^F_{fg} = \theta$.*

4.4 Macroeconomic Bias of Technical Change

We can also leverage Proposition 7 to characterize the macroeconomic bias of technical change in the aggregate production function.

Proposition 10. *(Bias of Technical Change) The biases towards one factor vs. another of the different technology shocks in the aggregate production function are given by*

$$\frac{B^F_{fgj}}{1+B^F_{fgj}} = (I_{(f)} - I_{(g)})'(I-\Gamma)^{-1}\hat{\delta}_{(j)},$$

where Γ is the $F \times F$ factor share propagation matrix defined by

$$\Gamma_{hh'} = -\sum_{k \in 1+N} (\theta_k - 1)\lambda_k Cov_{\Omega^{(k)}}\left(\Psi_{(h')}, \Psi_{(h)}/\Lambda_h\right),$$

$\hat{\delta}$ is the $F \times 1+N$ factor share impulse matrix defined by

$$\hat{\delta}_{hj} = \sum_{k \in 1+N} (\theta_k - 1)\lambda_k Cov_{\Omega^{(k)}}\left(\Psi_{(j)}, \Psi_{(h)}/\Lambda_h\right),$$

$\hat{\delta}_{(j)}$ is its jth column, I is the $F \times F$ identity matrix, and $I_{(f)}$ and $I_{(g)}$are its fth and gth columns.

In Section 3.4, we showed that the bias B^C_{fgj} of technical change in the aggregate cost function was a weighted sum of the departure from one $\theta_k - 1$ of the microeconomic elasticities of substitution in production with weights given by sufficient statistics $\lambda_k Cov_{\Omega^{(k)}}(\Psi_{(j)}, \Psi_{(f)}/\Lambda_f - \Psi_{(g)}/\Lambda_g)$ of the input-output network. For the aggregate production function, B^F_{fgj} is determined by departures from one $\theta_k - 1$ of microeconomic elasticities of substitution and by sufficient statistics $\lambda_k Cov_{\Omega^{(k)}}(\Psi_{(j)}, \Psi_{(h)}/\Lambda_h)$

and $\lambda_k Cov_{\Omega^{(k)}}(\Psi_{(h')}, \Psi_{(h)}/\Lambda_h)$ of the input-output network. However, it is no longer a weighted sum of the departures from one of the microeconomic elasticities of substitution, and it depends on a longer list of input-output network sufficient statistics. In fact, B^F_{fgj} is now a *nonlinear* function of the sufficient statistics $(\theta_k - 1)\lambda_k Cov_{\Omega^{(k)}}(\Psi_{(j)}, \Psi_{(h)}/\Lambda_h)$ and $(\theta_k - 1)\lambda_k Cov_{\Omega^{(k)}}(\Psi_{(h')}, \Psi_{(h)}/\Lambda_h)$.

As in case of the aggregate cost function, the network-irrelevance result for MESs in the aggregate production function stated in Proposition 9 does not extend to the bias of technical change. In general, the network matters for the bias of technical change, even when all the microeconomic elasticities of substitution θ_k are identical. Once again, the Cobb-Douglas is the one case where it doesn't: when all the microeconomic elasticities θ_k are unitary so that $\theta_k = 1$, technical change in unbiased with $B^F_{fgj} = 0$ for all f, g, and j, no matter what the structure of the network is.

As already mentioned, in Sections 6 and 6.5, we will present some simple theoretical and quantitative examples to illustrate the results of Sections 3 and 4.. In Section 7, we will also generalize these results to non-nested-CES economies. Before doing so however, we now turn to the questions of factor aggregation and network factorization and relate our results to the Cambridge-Cambridge controversy.

5 Factor Aggregation, Network Factorization, and the Cambridge-Cambridge Controversy

> "The production function has been a powerful instrument of miseducation. The student of economic theory is taught to write $Y = F(K, L)$ where L is a quantity of labour, K a quantity of capital and Y a rate of output of commodities. He is instructed to [...] measure L in man-hours of labour; he is told something about the index-number problem involved in choosing a unit of output ; and then he is hurried on [...], in the hope that he will forget to ask in what units K is measured. Before ever he does ask, he has become a professor, and so sloppy habits of thought are handed on from one generation to the next." — Robinson (1953)

As described earlier, the Cambridge-Cambridge controversy was a decades-long debate about the foundations of the aggregate production function. The broader context of the controversy was a clash between two views of the origins of the returns to capital. The first one is the Marxist view of the return to capital as a rent determined by political economy and monopolization. The second one is the marginalist view of the competitive return to capital determined by technology, returns to scale, and scarcity. The marginalist view is encapsulated in the "three key parables" of neoclassical writers (Jevons, Bohm-Bawerk, Wicksell, Clark) identified by Samuelson (1966): (1) the rate of interest is determined by technology ($r = F_K$); (2) there are diminishing returns to capital (K/Y and K/L are decreasing in r); and (3) the distribution of income is determined by relative factor scarcity (r/w is decreasing in K/L). These parables are consequences of having a per-period neoclassical aggregate production function $F(K, L)$ which has decreasing returns in each of its arguments.

In his famous "Summing Up" QJE paper (Samuelson, 1966), Samuelson, speaking for the Cambridge US camp, finally conceded to the Cambridge UK camp and admitted that indeed, capital could not be aggregated. He produced an example of an economy with "re-switching" : an economy where, as the interest rate decreases, the economy switches from one technique to the other and then back to the original technique. This results in a non-monotonic relationship between the capital-labor ratio as a function of the rate of interest r.

Since the corresponding capital-labor and capital-output ratios are non-monotonic functions of the rate of interest, this economy violates the first two of the three key parables. It is impossible to represent the equilibrium of the economy with a simple neoclassical model with a neoclassical aggregate production function with capital and labor, and where output can be used for consumption and investment.

Importantly, this result is established using *valuations* to compute the value of the capital stock index as sum of the values of the existing vintages of techniques, i.e. the net-present-value of present and future payments to nonlabor net of the net-present-value of present and future investments. The value of the capital stock depends on the rate of interest. Basically, the physical interpretation of capital is lost when it is aggregated in this *financial* way, and so are basic technical properties such as decreasing returns.[16]

[16] A historical reason for the focus of the controversy on the aggregation of capital as opposed to labor was the view held by the participants there was a natural physical unit in which to measure labor, man-hours. This view rests on the debatable assumption that different forms of labor, such as skilled labor and unskilled labor for example are perfect substitutes. Another historical reason was that some participants in the controversy took the view that labor could be reallocated efficiently across production units in response to shocks whereas capital was stuck in the short run, which they thought made the aggregation of capital more problematic. From the perspective of this paper, the aggregation problem for capital is not meaningfully different from that of labor. In general, outside of knife-edge cases, factors that are not perfectly substitutable or which cannot be reallocated cannot be aggregated.

The reactions to the Cambridge-Cambridge controversy were diverse. Post-Keynesians, like Pasinetti, considered neoclassical theory to have been "shattered" by their critiques.[17] Samuelson (and others like Franklin Fisher) on the other hand became invested in the view that one should develop disaggregated models of production. For example, Samuelson concluded his "A Summing Up" paper with this:

> "Pathology illuminates healthy physiology [...] If this causes headaches for those nostalgic for the old time parables of neoclassical writing, we must remind ourselves that scholars are not born to live an easy existence." — Samuelson (1966).

Solow was more ambivalent:

> "There is a highbrow answer to this question and a lowbrow one. The highbrow answer is that the theory of capital is after all just a part of the fundamentally microeconomic theory of the allocation of resources, necessary to allow for the fact that commodities can be transformed into other commodities over time. Just as the theory of resource allocation has as its 'dual' a theory of competitive pricing, so the theory of capital has as its 'dual' a theory of intertemporal pricing involving rentals, interest rates, present values and the like. The lowbrow answer, I suppose, is that theory is supposed to help us understand real problems, and the problems that cannot be understood without capital-theoretic notions are those connected with saving and investment. Therefore the proper scope of capital theory is the elucidation of the causes and consequences of acts of saving and investment. Where the highbrow approach tends to be technical, disaggregated, and exact, the lowbrow view tends to be pecuniary, aggregative, and approximate. A middlebrow like myself sees virtue in each of these ways of looking at capital theory. I am personally attracted by what I have described as the lowbrow view of the function of capital theory. But as so often happens, I think the highbrow view offers indispensable help in achieving the lowbrow objective." — Solow (1963).

In the mid 60s, an "MIT school" arose, which attempted to make progress on the study of disaggregated models of production. It impact was limited. Of course it didn't help that the re-switching example that concluded the Cambridge-Cambridge controversy seemed

[17] See for example Pasinetti et al. (2003).

so exotic. Later, developments in the growth literature, the arrival of real business cycle models, and the rational expectations revolution shifted the mainstream of the profession (with a few notable exceptions) away from these questions of heterogeneity and aggregation and towards dynamics and expectations.

In our opinion, the general neglect of these questions is unfortunate, and we hope that our work will contribute to reviving interest in these important topics. This section can be seen as a historical detour to make contact with the issues that preoccupied the protagonists of the Cambridge-Cambridge controversy. First, in Section 5.1, we show that generically, capital, or for that matter, any group of distinct factors, cannot be *physically* aggregated. Second, in Section 5.2, we give useful sufficient conditions for the possibility of *physically* factorizing the production network into components which can be represented via a sub-aggregate production functions. Third, in Section 5.3, we show how to capture Samuelson's reswitching example showing that capital cannot be linearly aggregated *financially* with *valuations* using our formalism.

The general lesson from this section is that the details of the production network matter, that outside of very knife-edge special cases, aggregating factors violates the structure of the network, and hence that it also changes the properties of the model. As a result, attempting to capture a disaggregated model of production by directly postulating an aggregate model does not work outside of very special cases.

5.1 Factor Aggregation

We study the aggregate production and cost functions of an economy with more than three factors. For brevity, we only treat the case of the aggregate cost function in nested-CES economies. The analysis of the general non-nested-CES case is similar, using the generalizations presented in Section 7. Similar proofs can be given in the case of the aggregate production function.[18] We also abstract from productivity shocks in our discussion (hold them fixed), but similar reasoning can be extended to productivity shocks.

Consider a non-trivial partition $\{F_i\}_{i \in I}$ of the set factors F, i.e. such that there exists an element of the partition comprising strictly more than 1 and strictly less than F factors. We say factors can be aggregated according in the partition $\{F_i\}_{i \in I}$ if there exists a set of

[18]The results also extend the "hybrid" case of an economy where some factors are in inelastic supply and some factors are in perfectly elastic supply, as in the steady state of a Ramsey model.

functions $\tilde{C}$ and $\tilde{g}_i$ such that

$$C(w_1, \ldots, w_F, Y) = \tilde{C}(g_1(\{w_f\}_{f \in F_1}), \ldots, g_I(\{w_f\}_{f \in F_I}), Y).$$

Similarly, we say that factors can be aggregated in to the partition $\{F_i\}_{i \in I}$, up to an nth order approximation, if there exists a set of functions $\tilde{C}$ and $\tilde{g}_i$ such that for all $m \leq n$ and $(f_1, f_2, \cdots, f_m) \in F^m$,

$$\frac{d^m \log C(w_1, \ldots, w_F, Y)}{dw_{f_1} \cdots dw_{f_m}} = \frac{d^m \log \tilde{C}(\tilde{g}_1(\{w_f\}_{f \in F_1}), \ldots, \tilde{g}_I(\{w_f\}_{i \in F_I}), Y)}{dw_{f_1} \cdots dw_{f_m}}.$$

In words, the factors can be aggregated up to the nth order, if there exists a separable function whose derivatives coincide with C up to the nth order.

A strict subset F_i of factors can always be aggregated locally to the first order by matching the shares of these factors in revenue.[19] But this aggregation fails to the second order, and by implication, it also fails globally. Indeed, and abstracting from productivity shocks, by the Leontief-Sono theorem, the strict subset F_i of factors can be globally aggregated in the aggregate cost function if and only if C_{w_f}/C_{w_g}, is independent of w_h for all $(f, g) \in F_i^2$ and $h \in F - F_i$. This is equivalent to the condition that

$$\frac{d^2 \log C}{d \log w_h d \log w_f} - \frac{d^2 \log C}{d \log w_h d \log w_g} = 0$$

or equivalently that $\sigma^C_{hf} = \sigma^C_{hg}$ for all $(f, g) \in F_i^2$, $h \in F - F_i$, and vector of factor prices. Using Proposition 2, this equation can be rewritten as

$$\sum_{k \in 1+N} (\theta_k - 1) \lambda_k Cov_{\Omega^{(k)}}(\Psi_{(h)}, \Psi_{(f)}/\Lambda_f - \Psi_{(g)}/\Lambda_g) = 0.$$

It is clear that this property is not generic: starting with an economy where this property holds, it is possible to slightly perturb the economy and make it fail. Indeed, suppose that the property holds at the original economy for a given vector of factor prices. Consider a set F_i an element of the partition comprising strictly more than 1 and strictly less than F factors. If $Cov_{\Omega^{(k)}}(\Psi_{(h)}, \Psi_{(f)}/\Lambda_f - \Psi_{(g)}/\Lambda_g) \neq 0$ for some k, $(f, g) \in F_i^2$, and $h \in F - F_i$, then it is enough to perturb the elasticity θ_k to make the property fail. If

[19] A loglinear approximation of the aggregate cost function is trivially separable in every partition, and is a first-order approximation. By Proposition 1, the log-linear approximation sets the elasticity of the aggregate cost function with respect to the wage of each factor equal to the revenue share of that factor.

$Cov_{\Omega^{(k)}}(\Psi_{(h)}, \Psi_{(f)}/\Lambda_f - \Psi_{(g)}/\Lambda_g) = 0$ for all k, $(f,g) \in F_i^2$, and $h \in F - F_i$, then we need to perturb the network to bring ourselves back to the previous case. It is enough to introduce a new producer producing only from factors h and f and selling only to final demand, with a small share ϵ, and scale down the other expenditure share in final demand by $1-\epsilon$. We can choose the exposures of the new producer to h and f such that $Cov_{\Omega^{(0)}}(\Psi_{(h)}, \Psi_{(f)}/\Lambda_f - \Psi_{(g)}/\Lambda_g) \neq 0$ for $\epsilon \neq 0$ small enough. This leads us the following proposition.

Proposition 11. *(Conditions for Factor Aggregation) Consider an economy with more than three factors and a non-trivial partition $\{F_i\}_{i \in I}$ of the set F of factors. In the aggregate production function, the factors can be aggregated in to the partition if and only if $\sigma_{hf}^F = \sigma_{hg}^F$ for all $i \in I$, $(f,g) \in F_i^2$, $h \in F - F_i$, and vector of factor supplies. Similarly, in the aggregate cost function, the factors can be aggregated according to the partition if and only if $\sigma_{hf}^C = \sigma_{hg}^C$ for all $i \in I$, $(f,g) \in F_i^2$, $h \in F - F_i$, and vector of factor prices. The conditions for factor aggregation according to a given partition in the aggregate production and cost functions are equivalent. Generically, these properties do not hold.*

The capital-aggregation theorem of Fisher (1965) can be seen through the lens of this proposition. It considers an economy with firms producing perfectly-substitutable goods using firm-specific capital and labor. It shows that the different capital stocks can be aggregated into a single capital index in the aggregate production function if and only if all the firms have the same production function up to a capital-efficiency term. In this case, and only in the case, all the MESs between the different firm-specific capital stocks and labor in the aggregate production function are all equal to each other, and are equal to the elasticity of substitution between capital and labor of the common firm production function.

Proposition 11 also shows that generically, capital, or indeed any other factor, cannot be aggregated. In other words, disaggregated production models cannot be avoided. Our approach in the previous sections acknowledges this reality and start with as many disaggregated factors as is necessary to describe technology. Our results take disaggregated models and seek to characterize their properties in terms of standard constructs such as the aggregate production and cost functions, marginal products of factors and factor demands, and elasticities of substitution between factors.

5.2 Network Factorization and Sub-Aggregate Production Functions

There is one frequently-occurring network structure under which we can establish a powerful form of network aggregation. This result can easily be conveyed at a high level of generality without requiring the economy to be of the nested-CES form. We need the following definition.

Definition. Let I be a subset of nodes. Let M be the set of nodes $j \notin I$ with $\Omega_{ij} \neq 0$ for some $i \in I$. We say (r, I, M) is an *island*, if

1. There is a unique node $r \in I$ such that $\Omega_{ir} = 0$ for every $i \in I$.

2. $\Omega_{ji} = 0$ for every $i \in I - \{r\}$ and every $j \notin I$.

3. $\Omega_{kj} = 0$ for every $j \in M$ and $k \notin I$.

We call r the export of the island. We say M are imports of the island. The imports of the island can be factors or non-factors. With some abuse of notation, we denote by x_{ri} the total imports of good $i \in M$ of the whole island (r, I, M). In the case where $f \in M$ is a factor, we also use the notation L_{rf}. See Figure 1 for a graphical illustration.

Note that the requirement that the export r of the island not be used as an intermediate input by other producers in the island is merely a representation convention: if it is not the case, we can always introduce a fictitious producer which transforms the good into an export using a one-to-one technology. The same remark applies to the requirement that imports of the island not be used by producers outside of the island: if a particular import is used by another producer outside of the island, we can always introduce a fictitious producer which transforms the good into an import using a one-to-one technology.

Given an island (r, I, M), we can define an associated island sub-aggregate production function with the island's imports as factors and its exports as aggregate output:

$$F_r\left(\{x_{ri}\}_{i \in M}, \{A_i\}_{i \in I}\right) = \max y_r, \tag{11}$$

subject to

$$y_j = A_j F_j(\{x_{jk}\}_{k \in I - \{r\} + M}) \quad (j \in I),$$

$$\sum_{i \in I} x_{ij} = x_{rj} \quad (j \in I - \{r\} + M).$$

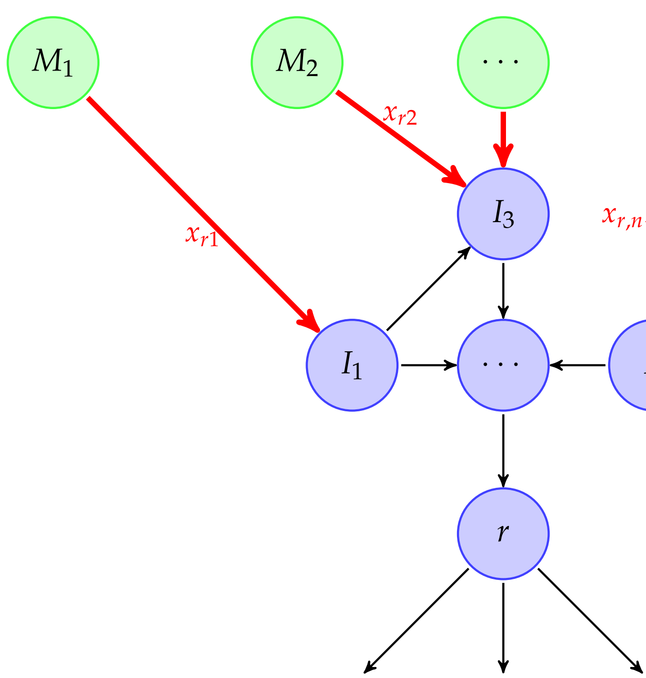

Figure 1: Illustration of an island (r, I, M) within a broader network. The nodes in the island I are in blue, the imports M are in green, and the export of the island is denoted by r. The figure only shows the island, its imports, and its export. This island is embedded in a broader network which is not explicitly represented in the figure.

With some abuse of notation, we use the same symbol F_r to denote the endogenous island sub-aggregate production function that we have used to denote the exogenous production function of producer r. The arguments of the latter are the intermediate inputs used by producer r while those of the former are the imports of the island and the productivities of the different producers in the island. The island sub-aggregate production function can be characterized using the same methods that we have employed for the economy-wide aggregate production function throughout the paper.

The planning problem defining the economy-wide aggregate production function can then be rewritten by replacing all the nodes in the island by its sub-aggregate production function:

$$F(L_1, \ldots, L_N, A_1, \ldots, A_N) = \max \mathcal{D}_0(c_1, \ldots, c_N)$$

subject to

$$y_i = A_i F_i(\{x_{ij}\}_{j \in N-I+\{r\}+F}) \quad (i \in N - I),$$

$$y_r = F_r \left(\{x_{ri}\}_{i \in M}, \{A_i\}_{i \in I} \right),$$

$$c_i + \sum_{j \in N-I+\{r\}} x_{ji} = y_i \quad (i \in N - I + \{r\}),$$

$$\sum_{i \in N-I+\{r\}} x_{if} = L_f \quad (f \in F),$$

where F_r is the island sub-aggregate production function.

So, if the economy contains islands, then the economy-wide aggregate production function can be derived in two stages: by first solving the island component planning problems (11) giving rise to the island sub-aggregate production functions, and then by solving the economy-wide problem giving rise to the economy-wide aggregate production function which uses the island sub-aggregate production functions. To describe this recursive structure, we say that the production network has been factorized.

Proposition 12 (Network Factorization)**.** *Let (r, I, M) denote an island. Then the economy-wide aggregate production function depends only on $\{A_i\}_{i \in I}$ and $\{x_{ri}\}_{i \in M} = \{y_i\}_{i \in M}$ via the island aggregate production function $F_r\left(\{x_{ri}\}_{i \in M}, \{A_i\}\right)$. In particular, if all the imports of the island are factors so that $M \subseteq F$, then the factors can be aggregated according to the partition $\{M, F - M\}$.*[20]

The factor-aggregation theorems of Fisher (1982) and Fisher (1983) correspond to special cases of the conditions in the second part of this proposition. The first part of the proposition is particularly useful in disaggregated intertemporal models. In some cases, intertemporal linkages can be represented via a set of capital stocks and their laws of motions via production functions with investment as inputs. This is the case, for example, of the post-Keynesian reswitching model studied in the next section. Even though will not pursue this particular representation of the example, it will prove helpful in understanding some of the obstacles preventing the aggregation of this model into a simple one-good neoclassical growth model.

5.3 Re-Switching Revisited

We now turn to the post-Keynesian reswitching example in Samuelson (1966). Samuelson's example features an economy with two goods in every period: labor and output. Labor is in unit supply. Output is used for consumption, labor can be used to produce output using two different production functions (called "techniques"). The first technique

[20] In this case, we can write $F(L_1, \ldots, L_N, A_1, \ldots, A_N) = \tilde{F}\left(F_r\left(\left\{L_f\right\}_{f \in M}, \{A_i\}_{i \in I}\right), \{A_i\}_{i \notin I}, \{L_i\}_{i \notin M}\right)$.

combines 2 units of labor at $t-2$ and 6 units of labor at t to produce one unit of output at t. The second technique uses 7 units of labor at $t-1$ to produce one unit of output at t. Both techniques are assumed to have constant-returns-to-scale.

We focus on the steady state of this economy, taking the gross interest rate $R = 1 + r$ as given, where r is the net interest rate. The interest rate R is varied by changing the rate of time preferences $\beta = 1/R$ of the agent. By comparing the unit costs of production, it is easy to see that the second technique dominates for high and low values of the interest rate, and that the first technique dominates for intermediate values of the interest rate. Indeed, at a gross interest rate of one (a net interest rate of zero), the second technique is preferred because it has a lower total labor requirement (7 vs. 8); and at a high interest rate, the second technique is preferred because the two-period delay in production of the first technique is too costly. Therefore, the economy features reswitching: as the interest rate is increased, it switches from the second to the first technique and then switches back to the second technique.

This post-Keynesian example can be obtained as a limit of the sort of nested-CES economies that we consider, provided that we use the Arrow-Debreu formalism of indexing goods and factors by dates and to think about capital stocks as intermediate goods. The corresponding production network is represented in Figure 2. The diagram shows how in different periods, labor can be combined with intermediate goods produced from past labor to produce new intermediate and final goods.

In Figure 3 we plot some steady-values for this economy as a function of the interest rate. The capital-labor and capital-output ratios are non-monotonic functions of the interest rate, where the aggregate capital stock is computed via financial valuations as the net-present-value of the payments to capital or equivalently at its replacement cost. Using the notation in Figure 2, this means that the aggregate capital stock in period t is computed as $K_t = p_{x_{1,t}} x_{1,t} + p_{y_{2,t}} y_{2,t} = 2w_{t-2} x_{1,t} + 7w_{t-1} y_{2,t}$.

We could alternatively represent this economy using *per-period* production functions for consumption and for investments, where the factors would be the capital stocks corresponding to the quantities of the different vintages of the two techniques and labor, and the laws of motions for the different capital stocks would combine previous capital stocks and investments to produce new capital stocks. These different production functions would correspond to a factorization of the production network into separate islands.

For our purposes here, it is more convenient instead to work directly with the disaggregated economy in Arrow-Debreu intertemporal form, which is characterized by the

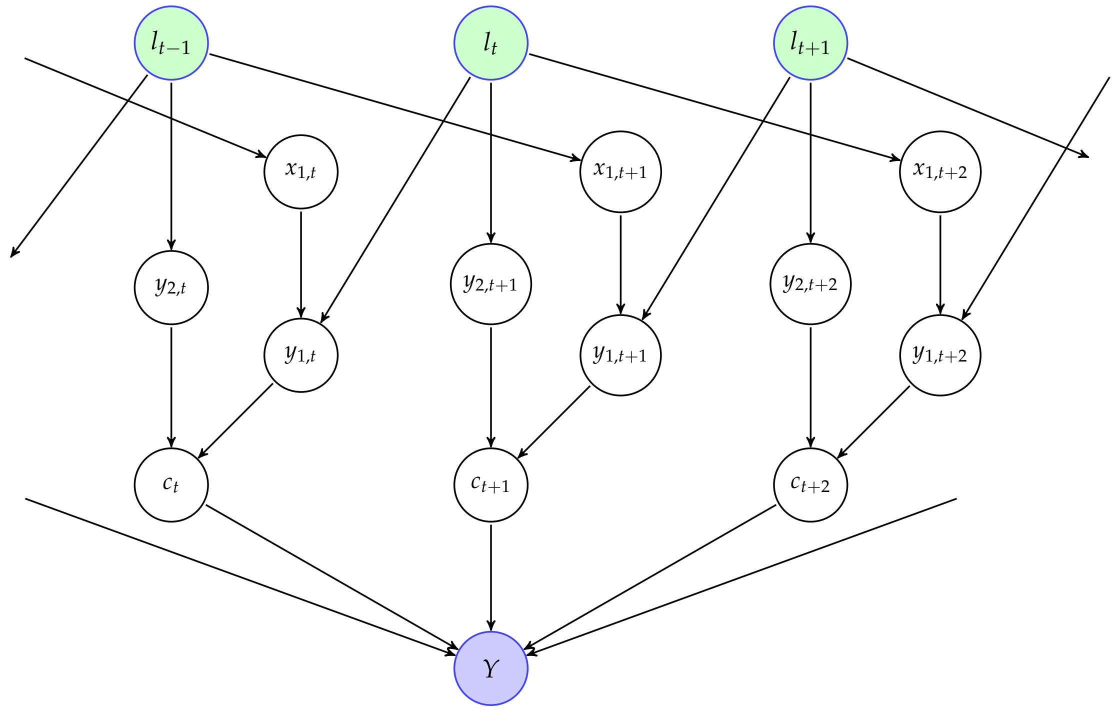

Figure 2: The production network underlying Samuelson's reswitching example. The arrows indicate the flow of goods. The green nodes are primary factors, and Y is aggregate output in this economy, which is perfectly substitutable across consumption units at different dates.

propagation equations (8) and (9). These equations are associated with an *intertemporal* production function characterizing the production of an intertemporal aggregate of consumption goods in all dates as a function labor in all dates.

For convenience, we consider a smoother version of the post-Keynesian example by imaging that the two techniques produce different goods which enter in consumption via a CES aggregator with a finite elasticity of substitution. The example obtains in the limit when that elasticity of substitution goes to infinity. As can be seen in Figure 4, the properties of the smoothed-out example resemble those of the original example.[21]

Now consider a simple one-good neoclassical growth model with a per-period neoclassical aggregate production function with capital and labor as its two arguments, where

[21]Marglin (1984) shows that re-switching cannot occur with smooth substitution in the sense that the composition of the basket of inputs used to produce each good cannot be the same for all goods for two different interest rates. One lesson of the original example is lost: that for no ranking of the two techniques in terms of "mechanization" or capital intensity is the economy necessarily becoming more mechanized or capital intensive as the interest rate decreases. Nonetheless, the lesson that the capital-labor and capital-output ratios are non-monotonic functions of the interest rate survives.

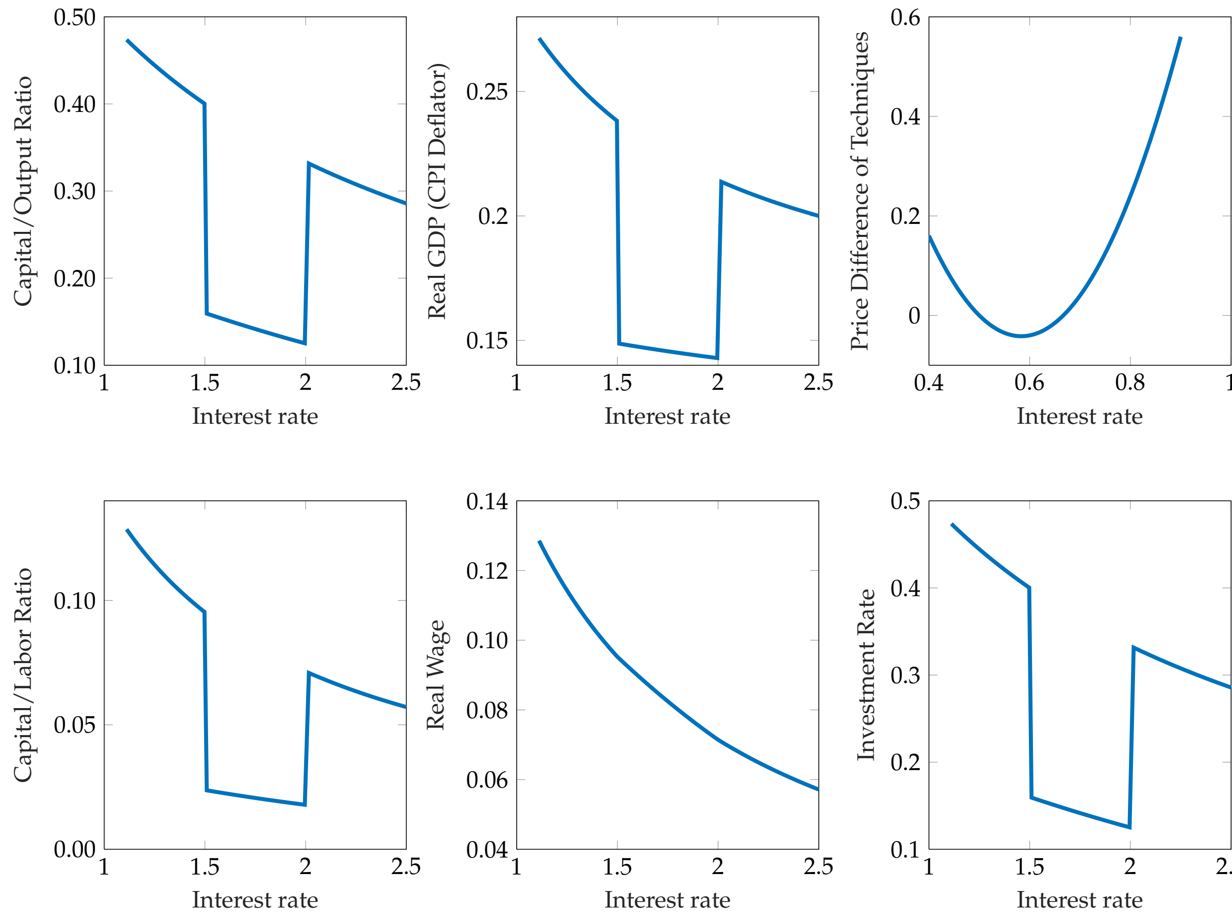

Figure 3: Samuelson's reswitching example

output can be used for investment and for consumption. In such a model, the homogeneous capital stock can be computed via financial valuations or equivalently at its replacement cost.

The question we now ask is whether we could represent the disaggregated post-Keynesian example as a version of the simple neoclassical model with an aggregate capital stock given by the sum of the values of the heterogeneous capital stocks in the disaggregated post-Keynesian example. The non-monotonicity of the capital-labor and capital-output ratios as a function of the interest rate shows that this is not possible. The simple neoclassical model could match the investment share, the capital share, the value of capital, and the value of the capital-output and capital-labor ratios of the original steady state of the disaggregated model, but not across steady states associated with different values

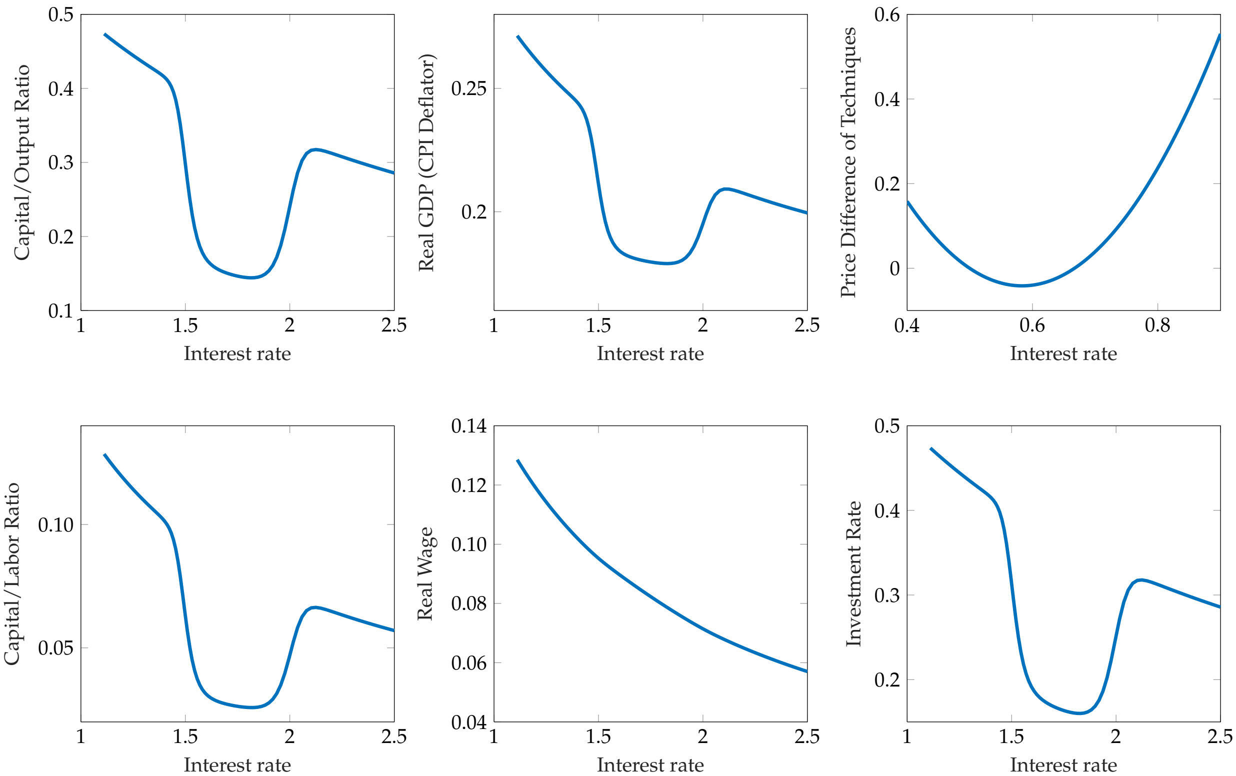

Figure 4: Samuelson's smoothed-out reswitching example

of the interest rate. In other words, aggregation via financial valuation fails.[22]

The production network of the simple one-good neoclassical growth economy is repre-

[22] Relatedly, consider a model with several capital stocks K_f and labor L. Each of the capital stocks depreciates from one period to the next with depreciation rate δ_{K_f} can be augmented by investing the final good with productivity $1/p_{K_f}$. The user cost of each capital stock is therefore $p_f(r+\delta_{K_f})$. To capture a steady state, we assume that the different capital stocks are in infinitely-elastic supply at these user costs, but that labor is in fixed supply. Appendix A.2 generalizes our equations to such a hybrid case with both factors in infinitely-elastic supply and factors in fixed supply.

We can compute the financial value of the aggregate capital stock $K = \sum_f p_{K_f} K_f$ in the model. We have

$$\frac{\mathrm{d}\log\left(\frac{\sum_f p_{K_f} K_f}{Y}\right)}{\mathrm{d}\log r} = -\left(\sum_f \frac{p_{K_f} K_f}{\sum_g p_{K_g} K_g} \frac{r}{r+\delta_{K_f}}\right)$$
$$\times \sum_f \frac{p_{K_f} K_f \frac{r}{r+\delta_{K_f}}}{\sum_g p_{K_g} K_g \frac{r}{r+\delta_{K_g}}} \left(1 + \sum_j \lambda_j(\theta_j - 1) Cov_{\Omega^{(j)}}\left(\sum_g \Lambda_{K_g} \frac{r+\delta_{K_f}}{r+\delta_{K_g}} \left(\frac{\Psi_{(K_g)}}{\Lambda_{K_g}} - \frac{\Psi_{(L)}}{\Lambda_L}\right), \frac{\Psi_{(K_f)}}{\Lambda_{K_f}}\right)\right).$$

Knowledge of the aggregate capital stock $K = \sum_f p_{K_f} K_f$ is not enough to infer the elasticity of the capital output ratio K/Y to the interest rate r. Knowledge of the values of all of the disaggregated capital stocks is required. Aggregation via financial valuations fails.

sented in Figure 5. The nodes (y_t, l_t, k_t) form an island where y_t is the export of the island, and k_t is the import of the island. Hence the relative price of k_t and l_t depends only on the quantity of k_t and l_t, and the problem can be studied in isolation via a neoclassical production function. The production function of the post-Keynesian reswitching example is different and it cannot be factorized into the same islands and represented with the same production functions. In other words, physical aggregation also fails.

One way to frame the lesson more generally is that the details of the production network matter. Aggregating factors changes the production network, and hence aggregation changes the properties of the model.

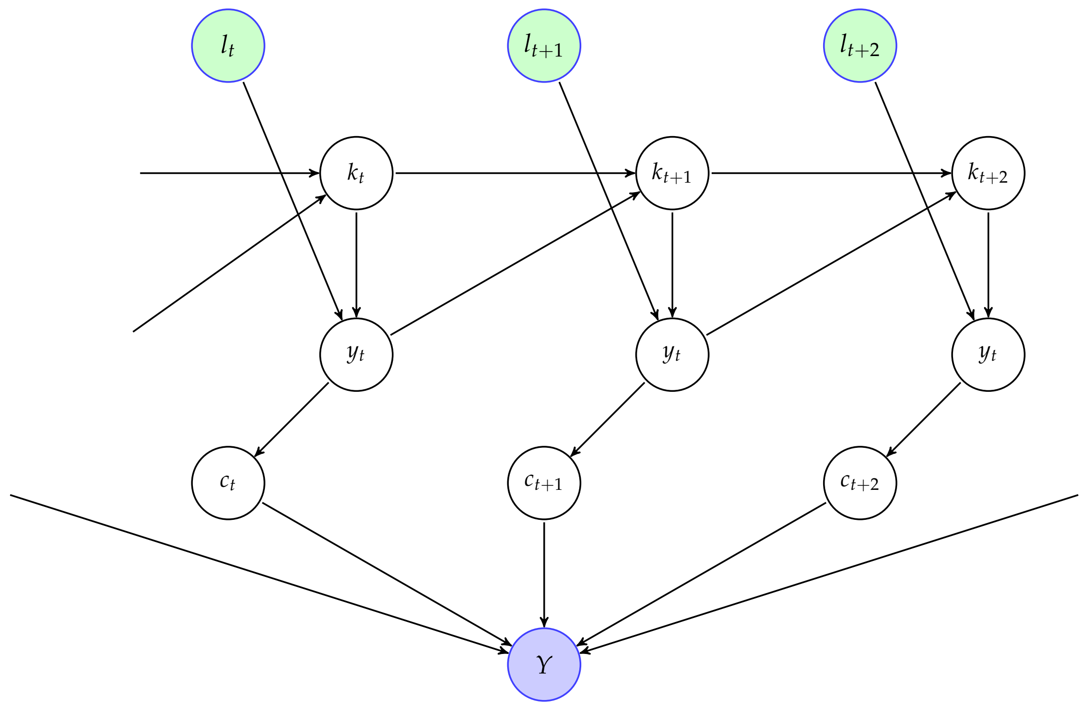

Figure 5: The production network underlying a simple one-good neoclassical growth model of an economy with a per-period production function with capital and labor as its two arguments, where output can be used for investment and for consumption. The arrows indicate the flow of goods. The green nodes are primary factors, and Y is aggregate output in this economy, which is perfectly substitutable across consumption units at different dates.

6 Simple Illustrative Examples

In this section, we provide four simple theoretical examples and a simple quantitative illustration. The first example is Hicksian in the sense that there are only two factors: the MESs in the aggregate production and cost functions are identical and are symmetric. The second example is non-Hicksian since it has three factors: the MESs in the aggregate production and cost functions are different and are asymmetric in general. The third example is the famous example of Houthakker (1955). The fourth example works out the macroeconomic bias of technical change which is capital augmenting at the microeconomic level in a disaggregated "task-based" model; it also shows that such a model can give rise to richer and more complex patterns than simpler models based on an aggregate production function in the sense that such technical change can be capital biased but not necessarily capital augmenting at the macroeconomic level. The fifth example is a simple quantitative illustration to capital-skill complementarity à la Griliches (1969) in the US economy, taking into account the multiplicity of sectors and their input-output linkages. We use the analysis to revisit the influential analysis in Krusell et al. (2000) of the role of these complementarities in the evolution over time of the skill premium.

6.1 A Hicksian Example with Two Factors

Our first example features two factors of production and producers with different factor intensities. A similar example is analyzed in Oberfield and Raval (2014), building on Sato (1975).[23] Each producer $1 \leq i \leq N$ produces from capital (K_i) and labor (L_i) according to

$$\frac{y_i}{\bar{y}_i} = \left(\omega_{iK} \left(\frac{K_i}{\bar{K}_i} \right)^{\frac{\theta_{iKL}-1}{\theta_{iKL}}} + \omega_{iL} \left(\frac{L_i}{\bar{L}_i} \right)^{\frac{\theta_{iKL}-1}{\theta_{iKL}}} \right)^{\frac{\theta_{iKL}}{\theta_{iKL}-1}}$$

and the final demand aggregator is

$$\frac{Y}{\bar{Y}} = \left(\sum_{i=1}^{N} \omega_{Di} \left(\frac{y_i}{\bar{y}_i} \right)^{\frac{\theta_D-1}{\theta_D}} \right)^{\frac{\theta_D}{\theta_D-1}}$$

[23] Sato (1975) only considered the case with two producers.

with $\omega_{iK} = 1 - \omega_{iL}$ and $\sum_{i=1}^{N} \omega_{Di} = 1$. Sales shares for goods and factors are given by $\lambda_D = 1$, $\lambda_i = \omega_{Di}$, $\Lambda_K = \sum_{i=1}^{N} \lambda_i \omega_{iK}$, and $\Lambda_L = \sum_{i=1}^{N} \lambda_i \omega_{iL}$.

For this example economy, the MES between capital and labor in the aggregate cost and production functions satisfy $\sigma^C_{LK} = \sigma^C_{KL} = \sigma^F_{LK} = \sigma^F_{KL} = \sigma_{LK}$, where σ_{LK} is given by

$$\sigma_{LK} = \sum_{i=1}^{N} \theta_{iKL} \lambda_i \frac{\omega_{iK}(1-\omega_{iK})}{\Lambda_K(1-\Lambda_K)} + \theta_D \frac{\sum_{i=1}^{N} \lambda_i (\omega_{iK} - \Lambda_K)^2}{\Lambda_K(1-\Lambda_K)},$$

where

$$\sum_{i=1}^{N} \lambda_i \frac{\omega_{iK}(1-\omega_{iK})}{\Lambda_K(1-\Lambda_K)} + \frac{\sum_{i=1}^{N} \lambda_i (\omega_{iK} - \Lambda_K)^2}{\Lambda_K(1-\Lambda_K)} = 1.$$

The MES between capital and labor σ_{LK} is a weighted average of the microeconomic elasticities of substitution between capital and labor θ_{iKL} and of the elasticity of substitution θ_D across producers in final demand. The weight on θ_{iKL} increases with its sales share λ_i and the heterogeneity in factor shares $\omega_{iK}(1-\omega_{iK})$ relative to the economy-wide heterogeneity in factor shares $\Lambda_K(1-\Lambda_K)$. It is zero when $\omega_{iK} = 0$ or $\omega_{iK} = 1$. The weight on θ_D increases in the heterogeneity in capital exposure across producers $\sum_{i=1}^{N} \lambda_i (\omega_{iK} - \Lambda_K)^2$ relative to the economy-wide heterogeneity in factor exposures $\Lambda_K(1-\Lambda_K)$. It is zero when $\omega_{iK} = \Lambda_K$ for all i.

6.2 A Non-Hicksian Example with Three Factors

Our second example extends the first example to include three factors. Each producer $1 \leq i \leq N$ produces from capital (K_i), skilled labor (H_i), and unskilled labor (L_i), according to

$$\frac{y_i}{\bar{y}_i} = \left(\omega_{iKH} \left(\omega_{iK} \left(\frac{K_i}{\bar{K}_i} \right)^{\frac{\theta_{iKH}-1}{\theta_{iKH}}} + \omega_{iH} \left(\frac{H_i}{\bar{H}_i} \right)^{\frac{\theta_{iKH}-1}{\theta_{iKH}}} \right)^{\frac{\theta_{iKH}}{\theta_{iKH}-1} \frac{\theta_{iKHL}-1}{\theta_{iKHL}}} + \omega_{iL} \left(\frac{L_i}{\bar{L}_i} \right)^{\frac{\theta_{iKHL}-1}{\theta_{iKHL}}} \right)^{\frac{\theta_{iKHL}-1}{\theta_{iKHL}}}$$

and the final demand aggregator is

$$\frac{Y}{\bar{Y}} = \left(\sum_{i=1}^{N} \omega_{Di} \left(\frac{y_i}{\bar{y}_i} \right)^{\frac{\theta_D - 1}{\theta_D}} \right)^{\frac{\theta_D}{\theta_D - 1}},$$

with $\omega_{iKH} = 1 - \omega_{iL}$, $\omega_{iK} = 1 - \omega_{iH}$, and $\sum_{i=1}^{N} \omega_{Di} = 1$. This economy can be written in normal form by introducing fictitious producers indexed by iKH producing a bundle of capital and skilled labor to be used as an input by producer i. Sales shares for goods and factors are given by $\lambda_D = 1$, $\lambda_i = \omega_{Di}$, $\lambda_{iKH} = \lambda_i \omega_{iKH}$, $\Lambda_K = \sum_{i=1}^{N} \lambda_i \omega_{iKH} \omega_{iK}$, $\Lambda_H = \sum_{i=1}^{N} \lambda_i \omega_{iKH} \omega_{iH}$, and $\Lambda_L = \sum_{i=1}^{N} \lambda_i \omega_{iL}$.

We start with the MESs in the aggregate cost function. For the sake of illustration, we focus on the MESs σ^C_{LK} and σ^C_{KL}. We have

$$\begin{aligned}\sigma^C_{LK} = \theta_D \sum_{i=1}^{N} \lambda_i \omega_{iKH} \omega_{iK} \left(\frac{\omega_{iKH}\omega_{iK}}{\Lambda_K} - \frac{\omega_{iL}}{\Lambda_L} \right) \\ + \sum_{i=1}^{N} \theta_{iKHL} \left[\lambda_i \omega_{iK} \frac{\omega_{iKH}\omega_{iK}}{\Lambda_K} + \lambda_i \omega_{iKH} \omega_{iK} \left(-\frac{\omega_{iKH}\omega_{iK}}{\Lambda_K} + \frac{\omega_{iL}}{\Lambda_L} \right) \right] \\ + \sum_{i=1}^{N} \theta_{iKH} \lambda_i \omega_{iH} \frac{\omega_{iKH}\omega_{iK}}{\Lambda_K},\end{aligned}$$

$$\begin{aligned}\sigma^C_{KL} = \theta_D \sum_{i=1}^{N} \lambda_i \omega_{iL} \left(\frac{\omega_{iL}}{\Lambda_L} - \frac{\omega_{iKH}\omega_{iK}}{\Lambda_K} \right) \\ + \sum_{i=1}^{N} \theta_{iKHL} \left[\lambda_i \frac{\omega_{iL}}{\Lambda_L} + \lambda_i \omega_{iL} \left(\frac{\omega_{iKH}\omega_{iK}}{\Lambda_K} - \frac{\omega_{iL}}{\Lambda_L} \right) \right].\end{aligned}$$

As is apparent from these formulas, in general, $\sigma^C_{LK} \neq \sigma^C_{KL}$. For instance, and by contrast with σ^C_{LK}, σ^C_{KL} does not depend on the microeconomic elasticities of substitution θ_{iKH} between capital and skilled labor. This follows from two observations: variations underlying the definition σ^C_{LK} vary w_K while keeping w_H and w_L constant while variations underlying the definition σ^C_{KL} vary w_L while keeping w_K and w_H constant; capital and skilled labor always enter in the CES nest iKH with elasticity θ_{iKH} while unskilled labor does not. In the special case where capital intensities are uniform across producers, σ^C_{LK} is independent of θ^D, and similarly, in the special case where labor intensities are uniform across producers, σ^C_{KL} is independent of θ^D. In general, and although verifying it requires some steps of algebra, both σ^C_{LK} and σ^C_{KL} are weighted averages of the microeconomic elasticities of substitution.

The expressions for the MESs in the aggregate production function σ^F_{KL} and σ^F_{LK} are more complex, and we omit them for brevity. Obtaining these equations requires solving a system of equations of two equations in two unknowns for the changes in factor shares $d\log\Lambda_L$ and $d\log\Lambda_K$ in response to a change $d\log K$ and $d\log L$ respectively (after having used the equation $\Lambda_K d\log\Lambda_K + \Lambda_H d\log\Lambda_H + \Lambda_L d\log\Lambda_L = 0$ to substitute out $d\log\Lambda_H$). In general, we have $\sigma^F_{LK} \neq \sigma^C_{KL}$, $\sigma^F_{LK} \neq \sigma^C_{LK}$, and $\sigma^F_{LK} \neq \sigma^F_{KL}$. Moreover, σ^F_{LK} and σ^F_{LK} are not weighted averages, or even linear functions, of the microeconomic elasticities of substitution.

These expressions simplify drastically in the case where factor intensities and microeconomic elasticities of substitution are uniform across producers so that ω_{iK}, ω_{iH}, ω_{iL}, ω_{iKH}, θ_{iKHL}, and θ_{iKH} are independent of i. In this case, the aggregate production and cost functions are of the nested-CES form. In particular, we get $\sigma^C_{KL} = \sigma^F_{KL} = \theta_{KHL}$, $\sigma^C_{LK} = \theta_{KHL}\Lambda_K/(\Lambda_K+\Lambda_H) + \theta_{KH}\Lambda_H/(\Lambda_K+\Lambda_H)$, and $\sigma^F_{LK} = (\Lambda_K+\Lambda_H)/(\Lambda_K/\theta_{KHL} + \Lambda_H/\theta_{KH})$. Hence we see that in this simple case, σ^C_{LK} and σ^F_{LK} are respectively the arithmetic and harmonic averages of the microeconomic elasticities θ_{KHL} and θ_{KH} with weights $\Lambda_K/(\Lambda_K+\Lambda_H)$ and $\Lambda_H/(\Lambda_K+\Lambda_H)$ and are therefore different in general.

6.3 Houthakker (1955)

Houthakker (1955) described how a disaggregated economy with fixed proportions and decreasing returns at the microeconomic level could give rise to a Cobb-Douglas aggregate production function with decreasing returns when the distribution of technical requirements across producers is a double Pareto. The model illustrates a divorce between microeconomic elasticities of substitutions between factors equal to 0, and macroeconomic elasticities of substitutions between factors equal to 1. The model is a particular limit case of our general model. In this section, we explain how to capture it using our formalism.[24]

The model features individual cells. Each individual j cell can produce up to ϕ_j units of output, where each unit of output requires $a_{1,j}$ units of factor L_1 and $a_{2,j}$ units of factor L_2. Using output as the numeraire, the unit is active in equilibrium if $1 - a_{1,j}w_1 - a_{2,j}w_2 \geq 0$. The total capacity of cells for which $a_{1,j}$ lies between a_1 and $a_1 + da_1$ and for which $a_{2,j}$ lies between a_2 and $a_2 + da_2$ can be represented by $\phi(a_1, a_2)da_1da_2$, where ϕ is the input-output distribution for the set of cells concerned. Total output and total factor demand

[24]Levhari (1968) generalizes Houthakker (1955) by deriving distributions of technical requirements across producers for which the aggregate production function is CES rather than simply Cobb Douglas. Sato (1969) in turn generalizes Levhari (1968) by allowing for microeconomic production to be CES rather than simply Leontief. All these models are particular cases of our general model.

are then given by

$$
\begin{aligned}
Y &= \int_0^{1/w_1} \int_0^{(1+a_1 w_1)/w_2} \phi(a_1, a_2) da_1 da_2, \\
L_1 &= \int_0^{1/w_1} \int_0^{(1+a_1 w_1)/w_2} a_1 \phi(a_1, a_2) da_1 da_2, \\
L_2 &= \int_0^{1/w_1} \int_0^{(1+a_1 w_1)/w_2} a_2 \phi(a_1, a_2) da_1 da_2.
\end{aligned}
$$

The last two equations implicitly give w_1 and w_2 as functions of L_1 and L_2, and plugging these functions back into the first equation gives output $Y = F(L_1, L_2)$ as a function of L_1 and L_2, thereby describing the aggregate production function of this economy. Characterizing this production function is difficult, and so Houthakker focused on the special case where the distribution of unit requirements is double Pareto with $\phi(a_1, a_2) = A a_1^{\alpha_1 - 1} a_2^{\alpha_2 - 1}$. He showed that in this case, the production function is given by

$$
F(L_1, L_2) = \Theta L_1^{\frac{\alpha_1}{\alpha_1 + \alpha_2 + 1}} L_2^{\frac{\alpha_2}{\alpha_1 + \alpha_2 + 1}},
$$

where

$$
\Theta = (\alpha_1 + \alpha_2 + 1) \left(\frac{A B(\alpha_1 + 1, \alpha_2 + 1)}{\alpha_1^{\alpha_1 + 1} \alpha_2^{\alpha_2 + 1}} \right)^{\frac{1}{\alpha_1 + \alpha_2 + 1}},
$$

where B is the beta function given by $B(\alpha_1 + 1, \alpha_2 + 1) = \int_0^1 t^{\alpha_1} (1 - t)^{\alpha_2}$.

To capture the model using our formalism, one first has to introduce more factors, because of decreasing returns to scale at the micro level. Specifically, we assume that over and above the factors L_1 and L_2, there is a different fixed factor L_j in unit supply for each producer j. Producer j produces output according to a Leontief aggregate $\min\{l_1/a_1, l_2/a_2, \phi_j l_j\}$, where l_1 is its use of factor L_1, l_2 its use of factor L_2, and l_j its use of factor L_j. The outputs of the different producers are then aggregated using a CES aggregate. Houthakker's model obtains in the limit where the elasticity of substitution of this CES aggregator goes to infinity. In this limit, the wage of the fixed factor of producer j can be computed as $(1 - a_{1,j} w_1 - a_{2,j} w_2)^+$ so that payments to all factors exhaust the revenues of producer j.

It is possible to obtain Houthakker's formulas in the particular case where the distribution is double Pareto by specializing our general formulas, but the calculations are tedious and so we refrain from doing so. The reason for this difficulty is that there are many factors: the two nonfixed factors and all the fixed factors. Our formulas solve for all these changes in shares simultaneously as the solution of a large system of linear equations. In the double Pareto case, it is actually possible to sidestep this difficulty and to solve directly for the changes in the shares in the two non-fixed factors, which turn out to be zero. This means that while the individual changes in the shares accruing to the fixed factors are nonzero, their sum is zero. Since the changes in these individual shares is not of direct interest for the question at hand, the direct method is preferable. Indeed, it is straightforward to see that

$$
\begin{aligned}
Y &= \frac{(\alpha_1 + \alpha_2 + 1) AB(\alpha_1 + 1, \alpha_2 + 1)}{\alpha_1 \alpha_2 w_1^{\alpha_1} w_2^{\alpha_2}}, \\
L_1 &= \frac{AB(\alpha_1 + 1, \alpha_2 + 1)}{\alpha_2 w_1^{\alpha_1 + 1} w_2^{\alpha_2}}, \\
L_2 &= \frac{AB(\alpha_1 + 1, \alpha_2 + 1)}{\alpha_1 w_1^{\alpha_1} w_2^{\alpha_2 + 1}}.
\end{aligned}
$$

This immediately implies that

$$
\begin{aligned}
\frac{w_1 L_1}{Y} &= \frac{\alpha_1}{\alpha_1 + \alpha_2 + 1}, \\
\frac{w_2 L_2}{Y} &= \frac{\alpha_2}{\alpha_1 + \alpha_2 + 1}.
\end{aligned}
$$

This in turn immediately implies that $\sigma^F_{L_1,L_2} = \sigma^F_{L_2,L_1} = \sigma^C_{L_1,L_2} = \sigma^C_{L_2,L_1} = 1$ as well as Houthakker's result that $F(L_1, L_2) = \Theta L_1^{\alpha_1/(\alpha_1+\alpha_2)} L_2^{\alpha_2/(\alpha_1+\alpha_2)}$.

6.4 Capital-Biased Technical Change in a Task-Based Model

In this section, we consider an example taken from Baqaee and Farhi (2018) and inspired by Acemoglu and Restrepo (2019). We compute the bias of technical change and explain its dependence on the microeconomic pattern of sales shares, factor intensities, and microeconomic elasticities of substitution. We then show that in a "task-based" economy with disaggregated production, a possible consequence of capital-augmenting technical change and automation at the microeconomic level is a simultaneous decline in both the labor share of income and the real wage at the macroeconomic level. This cannot happen in a simpler economy with an aggregate production function with capital-augmenting technical since such technical change would always increase the real wage. In other words, technical change which is capital augmenting at the microeconomic level is capital biased but not capital augmenting at the macroeconomic level. The impact of technical change is therefore richer and more complex in models of disaggregated production.

Assume that each producer, associated to a "task", produces from capital and labor according to

$$\frac{y_i}{\bar{y}_i} = \left(\omega_{iL} \left(\frac{\tilde{L}_i}{\bar{\tilde{L}}_i} \right)^{\frac{\theta_{KL}-1}{\theta_{KL}}} + \omega_{iK} \left(\frac{\tilde{K}_i}{\bar{\tilde{K}}_i} \right)^{\frac{\theta_{KL}-1}{\theta_{KL}}} \right)^{\frac{\theta_{KL}}{\theta_{KL}-1}}$$

with

$$\tilde{K}_i = \frac{A_{iK}}{\bar{A}_{iK}} K_i \qquad \text{and} \qquad \tilde{L}_i = \frac{A_{iL}}{\bar{A}_{iL}} L_i$$

and $\omega_{iK} = 1 - \omega_{iL}$. The consumer values the output of these tasks according to a CES aggregator

$$\frac{Y}{\bar{Y}} = \left(\sum_{i=1}^{N} \omega_{Di} \left(\frac{y_i}{\bar{y}_i} \right)^{\frac{\theta_D-1}{\theta_D}} \right)^{\frac{\theta_D}{\theta_D-1}},$$

with $\sum_{i=1}^{N} \omega_{Di} = 1$. Sales shares for goods and factors are given by $\lambda_D = 1$, $\lambda_i = \omega_{Di}$, $\Lambda_K = \sum_{i=1}^{N} \lambda_i \omega_{iK}$, and $\Lambda_L = \sum_{i=1}^{N} \lambda_i \omega_{iL}$.

Capital-biased technical change is modeled as a shock $d \log A_{kK} > 0$. Using our formulas, we can characterize the responses of the labor share and of the wage. The biases towards K vs. L of this technology shock in the aggregate cost function and production functions are given by

$$B^C_{KLkK} = (\theta_{KL} - 1)\lambda_k \frac{\omega_{kK}}{\Lambda_K} \frac{\omega_{kL}}{\Lambda_L} + (\theta_D - 1)\lambda_k \frac{\omega_{kK}}{\Lambda_K} (1 - \frac{\omega_{kL}}{\Lambda_L})$$

and

$$\frac{B^F_{KLkK}}{1 + B^F_{KLkK}} = \frac{(\theta_{KL} - 1)\lambda_k \frac{\omega_{kK}}{\Lambda_K} \frac{\omega_{kL}}{\Lambda_L} + (\theta_D - 1)\lambda_k \frac{\omega_{kK}}{\Lambda_K} (1 - \frac{\omega_{kL}}{\Lambda_L})}{1 + (\theta_{KL} - 1) \sum_{i=1}^{N} \lambda_i \frac{\omega_{iK}}{\Lambda_K} \frac{\omega_{iL}}{\Lambda_L} + (\theta_D - 1) \frac{1}{\Lambda_L \Lambda_K} Var_\lambda(\omega_{(L)})}.$$

When there is a single task k so that $\lambda_k = 1$, $\omega_{kL} = \Lambda_L$, and $\omega_{kK} = \Lambda_k$, the aggregate production function is CES with elasticity of substitution θ_{KL} between capital and labor, and the biases in the aggregate cost and production functions coincide $B^C_{KLkL} = B^F_{KLkL} = \theta_{KL} - 1$.

In general however, these two biases are different $B^C_{KLkL} \neq B^F_{KLkL}$. The bias in the aggregate cost function B^C_{KLkL} is a linear function of the different microeconomic elasticities of substitution whereas the bias in the aggregate production function B^F_{KLkL} is a nonlinear function of these elasticities. However, the signs of the two biases are identical. A capital-augmenting shock to task k is more likely to be biased towards capital vs. labor at the macroeconomic level when: (i) capital and labor are substitutes at the microeconomic level with $\theta_{KL} > 1$; and (ii) tasks are substitutes with $\theta_D > 1$ and task k is more capital intensive than the average task with $\omega_{kL}/\Lambda_L < 1$, or tasks are complements with $\theta_D < 1$ and task k is more labor intensive than the average task with $\omega_{kL}/\Lambda_L > 1$. The intuition for (i) is straightforward: in response to a positive shock, producer k substitutes expenditure towards capital if $\theta_{KL} > 1$ and towards labor if $\theta_{KL} < 1$. The intuition for (ii) is the following: a positive shock reduces the price of task k; if $\theta_D > 1$, the household substitutes expenditure towards task k, resulting in the reallocation of factors towards task k, which increases the overall expenditure on capital if $\omega_{kL}/\Lambda_L < 1$ and reduces it otherwise; if $\theta_D < 1$, the household substitutes expenditure away from task k, resulting in the reallocation of factors away from task k, which increases the overall expenditure on capital if $\omega_{kL}/\Lambda_L > 1$ and reduces it otherwise.

We now turn our attention to the effect of technical change on the real wage, holding factor supplies constant. For simplicity, we focus on the case where final demand is Cobb Douglas across tasks with $\theta_D = 1$. We also assume that capital and labor are substitutes at the microeconomic level with $\theta_{KL} > 1$, so that a capital-augmenting shock to task k is biased towards capital vs. labor at the macroeconomic level, i.e. a positive shock increases the capital share and decreases the labor share. As we shall now see, the effect of such a shock on the real wage is ambiguous:[25,26]

$$\frac{d\log w_L}{d\log A_{kK}} = \lambda_k \omega_{kK} \frac{1 + (\theta_{KL} - 1)\sum_i \lambda_i \left(\frac{\omega_{iL}}{\Lambda_L} - \frac{\omega_{kL}}{\Lambda_L}\right)\frac{\omega_{iK}}{\Lambda_K}}{1 + (\theta_{KL} - 1)\sum_i \lambda_i \frac{\omega_{iL}}{\Lambda_L}\frac{\omega_{iK}}{\Lambda_K}}.$$

If task k is more labor intensive than the average task with $\omega_{kL}/\Lambda_L > 1$, and capital

[25]We can compute this as a function of the aggregate production function using

$$\frac{d\log w_L}{d\log A_{kK}} = \frac{d\log F}{d\log A_{kK}} + \frac{d^2\log F}{d\log A_{kK} d\log L} / \frac{d\log F}{d\log L} = \lambda_k \omega_{kK} + \frac{d\log \Lambda_L}{d\log A_{kK}}.$$

[26]In the general case where $\theta_D \neq 1$, we have

$$\frac{d\log w_L}{d\log A_{kK}} = \lambda_k \omega_{kK} \frac{1 + (\theta_{KL} - 1)\sum_i \lambda_i \left(\frac{\omega_{iL}}{\Lambda_L} - \frac{\omega_{kL}}{\Lambda_L}\right)\frac{\omega_{iK}}{\Lambda_K} + (\theta_D - 1)\left[\frac{1}{\Lambda_K \Lambda_L} Var_\lambda(\omega_{(L)}) + \frac{\omega_{kL}}{\Lambda_L} - 1\right]}{1 + (\theta_{KL} - 1)\sum_{i=1}^N \lambda_i \frac{\omega_{iL}}{\Lambda_L}\frac{\omega_{iK}}{\Lambda_K} + (\theta_D - 1)\frac{1}{\Lambda_L \Lambda_K} Var_\lambda(\omega_{(L)})}.$$

and labor are highly substitutable with a high-enough value of θ_{KL}, then the real wage falls in response to a positive shock. This is because as task k substitutes away from labor and towards capital, labor is reallocated to other tasks who use labor less intensively. This reallocation of labor reduces the marginal product of labor and hence the real wage. These patterns cannot be generated in a simpler economy with an aggregate production function with capital-augmenting technical change since such a shock always increases the marginal product of labor and hence the real wage.[27]

6.5 Quantitative Application: Capital-Skill Complementarity and the Skill Premium

In this section, we briefly summarize a simple quantitative application of the results in Sections 3 and 4. We refer the reader to Appendix B for the details of the analysis. We show how to use these results to study capital-skill complementarity à la Griliches (1969) in the US economy, taking into account the multiplicity of sectors and their input-output linkages. We use the analysis to revisit the influential analysis in Krusell et al. (2000) of the role of these complementarities in the evolution over time of the skill premium.

Krusell et al. (2000) studies the relationship between the increasing skill premium and the rapid decline in the relative price of equipment investment goods. They find that complementarity between capital goods and high-skill labor can explain a large part of the increase in the skill premium. They use an aggregate model and directly postulate an exogenous aggregate production function. We revisit their analysis in the context of a disaggregated model where the aggregate production function emerges endogenously. We find that moving from the aggregate model to the disaggregated model reduces the MES between capital and high-skil labor from 0.67 to 0.66 and the MES between capital an low-skill labor from 1.05 to 0.93. These differences are enough to reduce by 20% the contribution of the decline in the relative price of equipment investment to the increase of the skill premium.

These particular results notwithstanding, we stress that our goal in this application is more to demonstrate quantitatively the dependence of macroeconomic elasticities of substitution on microeconomic primitives than to draw strong implications for the question of whether and how much complementarities between capital and skilled labor can explain the behavior of the skill premium.

[27] Indeed, suppose for example that there is a single task so that $\lambda_k = 1$. We get $\omega_{kL} = \Lambda_L$ and $\omega_{kK} = \Lambda_k$. This implies that $B^F_{KLkL} = \theta_{KL} - 1$ and $d \log w_L / d \log A_{kK} = \Lambda_k / \theta_{KL} > 0$. The result is true more generally.

7 Extensions

In Appendix A, we present a number of extensions. First, in Appendix A.1, we explain how to generalize our results to arbitrary economies beyond the nested-CES case with Hicks neutral technology shocks at the level of each nest. In Appendix A.3, we introduce alternative notions for the aggregate production and cost functions which separate final demand from technology and explain how to use our results to characterize these objects. In Appendix A.4, we explain how to generalize our results to economies where final demand is non-homothetic or where there are distortions.

8 Conclusion

This paper is part of a broader agenda to bring macro theory closer to micro data. As micro data becomes more plentiful, parsimonious reduced-form aggregate production functions look more antiquated. This paper takes a step towards realizing this goal by providing an organizing framework and some general characterizations of micro-founded aggregate production functions. The difference in this approach, compared to the classical analysis, is that it takes advantage of the information in granular micro-data and allows for data-intensive aggregation to arrive at the behavior of macroeconomic aggregates.

We have several other papers in this agenda. In Baqaee and Farhi (2019a), we focus on on the nonlinear macroeconomic impact on aggregate output of microeconomic shocks in efficient models. In Baqaee and Farhi (2019c), we extend our analysis to inefficient economies. In Baqaee and Farhi (2018), we characterize how shocks affect non-aggregate outcomes, namely how shocks propagate from one producer to another, and how microeconomic variables comove with one another in a production network, and we also show how to combine input-output production networks with heteorgeneous agents and non-homothetic final demand in closed economies. In Baqaee and Farhi (2019b), we show how to take these elements into account in trade models.

However, we think we have only scratched the surface of what is left to be done. Extending the analysis to allow for, amongst other things, entry, exit, and tractable stochastic dynamics, are further steps that need to be taken before theory can start to approximate reality. We are actively pursing these questions in ongoing research.

References

Acemoglu, D. and P. Restrepo (2019). Artificial intelligence, automation and work. In A. Goldfarb, J. Gans, and A. Agrawal (Eds.), *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics 109*(3), 659–684.

Atalay, E. (2017). How important are sectoral shocks? *American Economic Journal: Macroeconomics 9*(4), 254–80.

Baqaee, D. R. and E. Farhi (2018, June). Macroeconomics with Heterogeneous Agents and Input-Output Networks. NBER Working Papers 24684, National Bureau of Economic Research, Inc.

Baqaee, D. R. and E. Farhi (2019a). The macroeconomic impact of microeconomic shocks: Beyond Hulten's Theorem. *Econometrica (Forthcoming)*.

Baqaee, D. R. and E. Farhi (2019b). Networks, barriers, and trade.

Baqaee, D. R. and E. Farhi (2019c). Productivity and Misallocation in General Equilibrium. NBER Working Papers 24007, National Bureau of Economic Research, Inc.

Bewley, T. (1986). Stationary monetary equilibrium with a continuum of independently fluctuating consumers. *Contributions to mathematical economics in honor of Gérard Debreu 79*.

Blackorby, C. and R. R. Russell (1989). Will the real elasticity of substitution please stand up?(a comparison of the Allen/Uzawa and Morishima elasticities). *The American economic review 79*(4), 882–888.

Boehm, C. E., A. Flaaen, and N. Pandalai-Nayar (2019). Input linkages and the transmission of shocks: Firm-level evidence from the 2011 tōhoku earthquake. *Review of Economics and Statistics 101*(1), 60–75.

Cohen, A. J. and G. C. Harcourt (2003). Retrospectives: whatever happened to the cambridge capital theory controversies? *Journal of Economic Perspectives 17*(1), 199–214.

Felipe, J. and F. M. Fisher (2003). Aggregation in production functions: what applied economists should know. *Metroeconomica 54*(2-3), 208–262.

Fisher, F. M. (1965). Embodied technical change and the existence of an aggregate capital stock. *The Review of Economic Studies 32*(4), 263–288.

Fisher, F. M. (1982). Aggregate production functions revisited: the mobility of capital and the rigidity of thought. *The Review of Economic Studies 49*(4), 615–626.

Fisher, F. M. (1983). On the simultaneous existence of full and partial capital aggregates. *The Review of Economic Studies 50*(1), 197–208.

Griliches, Z. (1969). Capital-skill complementarity. *The review of Economics and Statistics*, 465–468.

Hamermesh, D. S. (1996). *Labor demand*. princeton University press.

Herrendorf, B., R. Rogerson, and A. Valentinyi (2013). Two perspectives on preferences and structural transformation. *American Economic Review 103*(7), 2752–89.

Hicks, J. (1932). *Theory of wages*. Springer.

Houthakker, H. S. (1955). The pareto distribution and the Cobb-Douglas production function in activity analysis. *The Review of Economic Studies 23*(1), 27–31.

Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of economic Dynamics and Control 17*(5-6), 953–969.

Hulten, C. and L. Nakamura (2017, April). Accounting for growth in the age of the internet: The importance of output-saving technical change. Working Paper 23315, National Bureau of Economic Research.

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

Johnson, G. E. (1997). Changes in earnings inequality: the role of demand shifts. *Journal of economic perspectives 11*(2), 41–54.

Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica 68*(5), 1029–1053.

Krusell, P. and A. A. Smith, Jr (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of political Economy 106*(5), 867–896.

Levhari, D. (1968). A note on houthakker's aggregate production function in a multifirm industry. *Econometrica: journal of the Econometric Society*, 151–154.

Marglin, S. A. (1984). *Growth, distribution, and prices*. Number 155. Harvard University Press.

McFadden, D. (1963). Constant elasticity of substitution production functions. *The Review of Economic Studies 30*(2), 73–83.

McKenzie, L. W. (1959). On the existence of general equilibrium for a competitive market. *Econometrica: journal of the Econometric Society*, 54–71.

Morishima, M. (1967). A few suggestions on the theory of elasticity. *Keizai Hyoron (Economic Review) 16*, 144–150.

Oberfield, E. and D. Raval (2014, September). Micro data and macro technology. Working Paper 20452, National Bureau of Economic Research.

Pasinetti, L. L., F. M. Fisher, J. Felipe, J. McCombie, and R. L. Greenfield (2003). Cambridge capital controversies. *The Journal of Economic Perspectives 17*(4), 227–232.

Robinson, J. (1953). The production function and the theory of capital. *The Review of Economic Studies 21*(2), 81–106.

Russell, R. R. (2017). Elasticities of substitution. Technical report.

Samuelson, P. A. (1966). A summing up. *The quarterly journal of economics 80*(4), 568–583.

Sato, K. (1969). Micro and macro constant-elasticity-of-substitution production functions in a multifirm industry. *Journal of Economic Theory 1*(4), 438–453.

Sato, K. (1975). *Production functions and aggregation*, Volume 90. North-Holland.

Sattinger, M. (1993). Assignment models of the distribution of earnings. *Journal of economic literature 31*(2), 831–880.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics 70*(1), 65–94.

Solow, R. M. (1963). *Capital theory and the rate of return*. North-Holland.

Stern, D. I. (2010). Derivation of the hicks, or direct, elasticity of substitution using the input distance function. *Economics letters 108*(3), 349–351.

A Appendix: Extensions

In this section, we present a number of extensions. First, in Section A.1, we explain how to generalize our results to arbitrary economies beyond the nested-CES case with Hicks neutral technology shocks at the level of each nest. In Section A.3, we introduce alternative notions for the aggregate production and cost functions which separate final demand from technology and explain how to use our results to characterize these objects. In Section A.4, we explain how to generalize our results to economies where final demand is non-homothetic or where there are distortions.

A.1 Beyond CES

In Sections 3 and 4, we confined our characterization of aggregate cost and production functions to the general class of nested-CES economies with Hicks-neutral technology shocks at the level of each nest.

Now each producer i has a production function $F_i(x_{i1}, \cdots, x_{iN}, L_{i1}, \cdots, L_{iF}, A_i)$ with associated cost function $C_i(p_1, \cdots, p_N, w_1, \cdots, w_F, A_i)$. Generically, we can normalize the technology shock A_i so that $d\log F/d\log A_i = d\log C/d\log A_i = 1$ at the point of interest.

We proceed in two successive steps. In the first step, we continue to assume that technology shocks are Hicks neutral at the producer level and explain how to deal with non-CES producers.[28] We then explain how to extend our results using the input-output substitution operator, which is a generalization of the input-output covariance operator defined in equation (7). In the second step, we relax the assumption of Hicks-neutral technical change at the producer level and explain how to deal with completely general biased technical change at the producer level.[29]

Dealing with Non-CES Producers

We start with the case where technical change is Hicks neutral at the producer level. We introduce the input-output substitution operator. We show how it can be used to gener-

[28] That technology shocks are Hicks neutral at the producer level means that we can write, with some abuse of notation, $F_i(x_{i1}, \cdots, x_{iN}, L_{i1}, \cdots, L_{iF}, A_i) = A_i F_i(x_{i1}, \cdots, x_{iN}, L_{i1}, \cdots, L_{iF})$ and $C_i(p_1, \cdots, p_N, w_1, \cdots, w_F, A_i) = A_i^{-1} C_i(p_1, \cdots, p_N, w_1, \cdots, w_F)$

[29] In contrast to our analysis of the nested-CES case, in this section, we purposefully eschew any relabelling of the network via the disaggregation of a producer into an network of producers (an island) or via the introduction of fictitious producers. Factor-biased technical change is modeled directly at the original producer level, where it is not necessarily factor-augmenting.

alize the results of Sections 3 and 4.

Definition. (Micreconomic Allen-Uzawa Elasticities of Substitution) For a producer k, let $\theta_k(l, l')$ denote the Allen-Uzawa elasticity of substitution in cost between inputs $l \neq l'$ in the cost of producer k:

$$\theta_k(l, l') = \frac{C_k d^2 C_k / (dp_l dp_{l'})}{(dC_k / dp_l)(dC_k / dp_{l'})} = \frac{\epsilon_k(l, l')}{\Omega_{kl'}},$$

where $\epsilon_k(l, l')$ is the elasticity of the demand by producer k for input l with respect to the price $p_{l'}$ of input l', and $\Omega_{kl'}$ is the expenditure share in cost of input l'.

This definition applies to good inputs and factor inputs using the aforementioned notation $p_{N+f} = w_f$. Because of the symmetry of partial derivatives, we have $\theta_k(l, l') = \theta_k(l', l)$. Because of the homogeneity of degree one of the cost function, we have the homogeneity identity $\sum_{l' \in N+F} \Omega_{kl'} \theta_k(l, l') = 0$.

Following Baqaee and Farhi (2019a), we introduce the *input-output substitution operator* for producer k:

$$\Phi_k(\Psi_{(i)}, \Psi_{(j)}) = -\sum_{(l,l') \in (N+F)^2} \Omega_{kl} [\delta_{ll'} + \Omega_{kl'}(\theta_k(l, l') - 1)] \Psi_{li} \Psi_{l'j}, \tag{12}$$

$$= \frac{1}{2} E_{\Omega^{(k)}} \left((\theta_k(l, l') - 1)(\Psi_i(l) - \Psi_i(l'))(\Psi_j(l) - \Psi_j(l')) \right), \tag{13}$$

where $\delta_{ll'}$ is the Kronecker symbol, $\Psi_i(l) = \Psi_{li}$, $\Psi_j(l') = \Psi_{l'j}$, and the expectation in the second line is over l and l'. The second line can be obtained from the first using the symmetry of Allen-Uzawa elasticities and the homogeneity identity.

When the production function of k is CES with elasticity of substitution θ_k, the cross Allen-Uzawa elasticities $\theta_k(l, l')$ are identical $\theta_k(l, l') = \theta_k$ for $l \neq l'$, the own Allen-Uzawa elasticities are given by $\theta_k(l, l) = -\theta_k(1 - \Omega_{kl}) / \Omega_{kl}$, and we recover

$$\Phi_k(\Psi_{(i)}, \Psi_{(j)}) = (\theta_k - 1) Cov_{\Omega^{(k)}}(\Psi_{(j)}, \Psi_{(i)}).$$

Even outside of the CES case, the input-output substitution operator $\Phi_k(\Psi_{(i)}, \Psi_{(j)})$ shares many properties with a covariance operator. For example, it is immediate to verify that: $\Phi_k(\Psi_{(i)}, \Psi_{(j)})$ is bilinear in $\Psi_{(i)}$ and $\Psi_{(j)}$; $\Phi_k(\Psi_{(i)}, \Psi_{(j)})$ is symmetric in $\Psi_{(i)}$ and $\Psi_{(j)}$; $\Phi_k(\Psi_{(i)}, \Psi_{(j)}) = 0$ whenever $\Psi_{(i)}$ or $\Psi_{(j)}$ is constant.

All of our results in Sections 3 and 4 can be generalized to non-nested-CES economies.

All that is needed is to replace terms of the form $(\theta_k - 1)Cov_{\Omega^{(k)}}(\Psi_{(j)}, \Psi_{(i)})$ by $\Phi_k(\Psi_{(i)}, \Psi_{(j)})$. For example, the result in Proposition 3 for the MES between factors in the aggregate cost function becomes

$$\sigma^C_{fg} - 1 = \sum_{k \in 1+N} \lambda_k \Phi_k(\Psi_{(g)}, \Psi_{(g)}/\Lambda_g - \Psi_{(f)}/\Lambda_f).$$

Just like in the nested-CES case, σ^C_{fg} is a weighted average of the microeconomic elasticities of substitution $\theta_k(l, l')$ and is equal to θ if the microeconomic elasticities of substitution are all equal to θ.

Intuitively, $\Phi_k(\Psi_{(i)}, \Psi_{(j)})$ captures the way in which k redirects expenditure share towards i in response to one percent change in the price of j. To see this, we make use of the following well-known result (see for example Russell, 2017): the elasticity of the expenditure share of producer k on input l with respect to the price of input l' is given by $\delta_{ll'} + \Omega_{kl'}(\theta_k(l, l') - 1)$.[30] Equation (12) says that the way k redirects expenditure share towards i in response to unit proportional decrease in the price of j depends on considering, for each pair of inputs l and l', how much the proportional decrease $\Psi_{l'j}$ in the price of l' induced by the decline in the price of j causes k to increase its expenditure share on l (measured by $-\Omega_{kl}[\delta_{ll'} + \Omega_{kl'}(\theta_k(l, l') - 1)\Psi_{l'j}$), and on the exposure of l to i (measured by Ψ_{li}).

Equation (13) says that the way k redirects expenditure share towards i in shares in response to a decline in the price of j depends on considering, for each pair of inputs l and l', whether or not increased exposure to j (measured by $\Psi_j(l) - \Psi_j(l')$), is aligned with increased exposure to i (measured by $\Psi_i(l) - \Psi_i(l')$), and whether l and l' are complements or substitutes (measured by $(\theta_k(l, l') - 1)$).

Dealing with Biased Technical Change at the Producer Level

We now further generalize the results of Sections 3 and 4 to the case where technical change is biased at the producer level. This generalization matters only for our results on the bias of technical change in the aggregate production and cost functions. It does not change anything to our results on the MESs between factors in the aggregate production and cost functions.

[30]This property is the reason we choose to use Allen-Uzawa elasticities at the producer level: because they easily give the elasticities of cost expenditure shares with respect to input prices. Morishima elasticities $\sigma^C_k(l, l') = \epsilon_k(l, l') - \epsilon_k(l', l')$ are better suited instead to give the elasticities of *relative* cost expenditure shares. Of course, the two concepts are related and so we could have used either one.

Definition. (Microeconomic Bias of Technical Change) For a producer k, we denote by b^C_{ijk} the *microeconomic bias* towards input i vs. j of a technology shock to producer k, defined by the elasticity of the ratio of the cost shares of inputs i vs. j of producer k with respect to a technology shock to producer k, holding input prices constant:

$$b^C_{ijk} = \frac{d\log(\Omega_{ki}/\Omega_{kj})}{d\log A_k}.$$

Similarly, we denote by b^C_{ik} the bias towards input i of a technology shock to producer k, defined by elasticity of the cost share of input i of producer k, holding input prices constant:

$$b^C_{ik} = \frac{d\log\Omega_{ki}}{d\log A_k} = \sum_{j\in N}\Omega_{kj}b^C_{ijk}.$$

Note that we have defined the bias of technical change towards inputs, be they factors or intermediate goods. These notions are purely microeconomic. They are defined at the level of an individual producer, and do not embed any general equilibrium whatsoever. In what follows, we will rely on the following properties: $\sum_{i\in N}\Omega_{ki}b^C_{ik} = 0$ and $\sum_{i\in N}\sum_{j\in N}\Omega_{ki}\Omega_{kj}b^C_{ijk} = 0$

In the presence of biased technical change at the producer level, the only difference to the analysis in Section A.1 is that we now must include new forcing terms in the propagation equations (5) and (6) in Proposition 2 and (8) and (9) in Proposition 7. More precisely the following term must be added on the right-hand sides of equations (5) and (8) for $d\log\lambda_i$:

$$\sum_{k\in 1+N}\sum_{j\in 1+N}\frac{\lambda_k}{\lambda_i}Cov_{\Omega^{(k)}}(b^C_{(k)},\Psi_{(i)})d\log A_k;$$

and the following term must be added on the right-hand sides of equations (6) and (9) for $d\log\Lambda_f$:

$$\sum_{k\in 1+N}\sum_{j\in 1+N}\frac{\lambda_k}{\Lambda_f}Cov_{\Omega^{(k)}}(b^C_{(k)},\Psi_{(f)})d\log A_k,$$

where $b^C_{(k)}$ is the k-th column of b^C_{jk}. These terms account for the direct and indirect effects through the network of the changes in input expenditures of each producer k due to the bias of its technical change $d\log A_k$.

Of course, these new terms do not appear in the results on the MESs between factors in the aggregate production and cost functions which do not depend on the nature of technical change. But they do enter the results on the bias of technical change in the aggregate production and cost functions. For example, the result in Proposition 5 for the bias of technical change in the aggregate cost function becomes

$$B^C_{fgj} \;=\; \lambda_j Cov_{\Omega^{(j)}}(b^C_{(j)}, \Psi_{(f)}/\Lambda_f - \Psi_{(g)}/\Lambda_g) + \sum_{k\in 1+N} \lambda_k \Phi_k(\Psi_{(j)}, \Psi_{(f)}/\Lambda_f - \Psi_{(g)}/\Lambda_g).$$

A.2 Hybrid Model with Infinitely Elastic and Inelastic Factors

We assume that some of the factors F_∞ are in infinitely-elastic supply and that the others F_0 are in fixed supply, with $F = F_\infty \cup F_0$. We then have

$$d\log\lambda_i = \sum_{k\in 1+N}(\theta_k - 1)\frac{\lambda_k}{\lambda_i} Cov_{\Omega^{(k)}}(\sum_{j\in N}\Psi_{(j)} d\log A_j + \sum_{g\in F}\Psi_{(g)} d\log L_g, \Psi_{(i)})$$
$$- \sum_{k\in 1+N}(\theta_k - 1)\frac{\lambda_k}{\Lambda_f} Cov_{\Omega^{(k)}}(\sum_{h\in F_0}\Psi_{(h)} d\log\Lambda_h - \sum_{h\in F_\infty}\Psi_{(h)}\,\mathrm{d}\log Y, \Psi_{(i)}),$$

where the elasticities of output and of factor shares solve the following system of linear equations

$$\mathrm{d}\log Y = \frac{1}{\sum_{f\in F_0}\Lambda_f}\left[\sum\sum_{j\in N}\lambda_j\,\mathrm{d}\log A_j - \sum_{g\in F_\infty}\Lambda_g\,\mathrm{d}\log w_g + \sum_{g\in F_0}\Lambda_g\,\mathrm{d}\log L_g - \sum_{g\in F_0}\Lambda_d\,\mathrm{d}\log\Lambda_g\right],$$

$$d\log\Lambda_f = \sum_{k\in 1+N}(\theta_k - 1)\frac{\lambda_k}{\Lambda_f} Cov_{\Omega^{(k)}}(\sum_{j\in N}\Psi_{(j)} d\log A_j + \sum_{g\in F}\Psi_{(g)} d\log L_g, \Psi_{(f)})$$
$$- \sum_{k\in 1+N}(\theta_k - 1)\frac{\lambda_k}{\Lambda_f} Cov_{\Omega^{(k)}}(\sum_{h\in F_0}\Psi_{(h)} d\log\Lambda_h - \sum_{h\in F_\infty}\Psi_{(h)}\,\mathrm{d}\log Y, \Psi_{(f)}).$$

A.3 Separating Technology from Final Demand

As mentioned before, in this paper, we have put ourselves under conditions where the existence of an aggregate output good can be taken for granted because final demand is homothetic. Given our definitions, the properties of the aggregate production and cost functions depend on final demand.

It is possible to define alternative notions which do not depend on final demand: the aggregate distance function and its associated aggregate cost function. These alternative notions must then be combined with final demand to derive general-equilibrium comparative static results. Our definitions intentionally take these two steps at once. We take this route because we want to put the focus on general-equilibrium comparative statics. However, our results can also be used to characterize the properties of aggregate distance and cost functions. This is because we have full flexibility in specifying final demand, meaning that the aggregate distance function and its associated cost function can be analyzed by specifying final demand to be Leontief (i.e. setting the elasticity of substitution between final goods to be zero).

To see this, we first define the aggregate distance function and the associated cost function. Define the aggregate technology set

$$\Gamma(A) = \{(L, c) : L \text{ can produce } c\},$$

where L is a vector of factors, c is a vector of final consumptions, and A is a vector of productivities. The aggregate distance function is defined as follows:

$$D(L, A, c) = \max_{\{\delta:(L/\delta, c)\in\Gamma(A)\}} \delta,$$

so that $(L, c) \in \Gamma(A)$ if and only if $D(L, A, c) \geq 1$. Similarly, and with some abuse of notation, we can define the aggregate cost function

$$C(w, c, A) = \min_{\{L:(L,c)\in\Gamma(A)\}} wL.$$

Whenever an output good $\mathcal{D}_0(c)$ exists because final demand is homothetic, the aggregate production function can be recovered from the aggregate distance function using

$$\frac{F(L, A)}{Y} = \max_{\{c:\mathcal{D}_0(c)\geq Y\}} D(L, A, c).$$

Going in the other direction is also possible. To capture an aggregate distance function via an aggregate production function, for a given vector of final output c, suppose that final demand has a Leontief form $\mathcal{D}_0 = \mathcal{D}_0^c$, where $\mathcal{D}_0^c(\hat{c}) = \min\{\hat{c}_1/c_1, \cdots, \hat{c}_N/c_N\}$, and then compute the associated aggregate production function F^c to recover

$$D(L, A, c) = \frac{F^c(L, A)}{\mathcal{D}_0^c(c)} = F^c(L, A).$$

Basically, the aggregate distance function is obtained by removing the possibility of substitution in final demand. Similar considerations apply to the relationship between the aggregate cost functions $C(w, Y, A)$ and $C(w, c, A)$. Therefore, and although we do not present them in this form, our results can be used to fully characterize the aggregate distance function $D(L, A, c)$ and the aggregate cost function $C(w, c, A)$.

A.4 Non-Homothetic Final Demand and Distortions

One advantage of the alternative notions defined in Section A.3 is that they can be applied to to economies that do not possess an "aggregate output" good. However, as already mentioned, these notions must then be combined with final demand to derive general-equilibrium comparative statics.

In Baqaee and Farhi (2018), we develop an alternative approach based on "propagation equations" which maintains the focus on general-equilibrium comparative static results. These equations generalize equations (5) and (6) in Proposition 2 and equations (8) and (9) in Proposition 7. These equations allow us to characterize the first- and second-order properties of real GDP when it is defined as a Divisa index (and does not correspond to any physical quantity) and hence to compute macroeconomic elasticities of substitution between factors and the macroeconomic bias of technical change along the same lines as in this paper. This methodology generalizes our results to environments with non-homothetic final demand. It also has the advantage of allowing us to also deal with economies with distortions in the same unified framework.

B Appendix: Quantitative Application in Section 6.5

We show how to use these results to study capital-skill complementarity à la Griliches (1969) in the US economy, taking into account the multiplicity of sectors and their input-output linkages. We use the analysis to revisit the influential analysis in Krusell et al. (2000) of the role of these complementarities in the evolution over time of the skill premium.

Krusell et al. (2000) studies the relationship between the increasing skill premium and the rapid decline in the relative price of equipment investment goods. They find that complementarity between capital goods and high-skill labor can explain a large part of the increase in the skill premium. They use an aggregate model and directly postulate an exogenous aggregate production function. We revisit their analysis in the context of a disaggregated model where the aggregate production function emerges endogenously. We find that moving from the aggregate model to the disaggregated model reduces the MES between capital and high-skil labor from 0.67 to 0.66 and the MES between capital an low-skill labor from 1.05 to 0.93. These differences are enough to reduce by 20% the contribution of the decline in the relative price of equipment investment to the increase of the skill premium. These particular results notwithstanding, we stress that our goal in this application is more to demonstrate quantitatively the dependence of macroeconomic elasticities of substitution on microeconomic primitives than to draw strong implications for the question of whether and how much complementarities between capital and skilled labor can explain the behavior of the skill premium.

Krusell et al. (2000) work with a nested CES aggregate production function of the form

$$\frac{Y}{\bar{Y}} = \left(\omega_L \left(\frac{X}{\bar{X}} \right)^{\frac{\theta_2 - 1}{\theta_2}} + (1 - \omega_L) \left(\frac{L}{\bar{L}} \right)^{\frac{\theta_2 - 1}{\theta_2}} \right)^{\frac{\theta_2}{\theta_2 - 1}},$$

with

$$\frac{X}{\bar{X}} = \left(\omega_H \left(\frac{H}{\bar{H}} \right)^{\frac{\theta_1 - 1}{\theta_1}} + \omega_K \left(\frac{K}{\bar{K}} \right)^{\frac{\theta_1 - 1}{\theta_1}} \right)^{\frac{\theta_1}{\theta_1 - 1}}.$$

Here L represents low-skill labor, H high-skill labor, and K is capital.

They estimate the model on macro data using the first-order conditions of the model and find an elasticity of substitution between skilled labor and capital to be $\theta_1 = 0.67$ and an elasticity of substitution between unskilled labor and the composite factor X to be $\theta_2 = 1.67$. They argue that these estimates are plausible by noting that they lie in the middle of the range of estimates in the microeconomic literature reported by Johnson (1997) and Hamermesh (1996). They then use their estimated model to perform a counterfactual and conclude that the decrease in the price of equipment investment goods can explain a large part of the increase in the skill premium over time.

We revisit their exercise in a calibrated disaggregated model with 66 sectors and input-output linkages. We consider a model with 5 distinct micro elasticities of substitution $(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$. The parameter θ_0 is the elasticity of substitution across industries in consumption, θ_1 is the elasticity of substitution between high-skilled labor and capital, θ_2 is the elasticity of substitution between low-skilled labor and a composite factor consisting of high skilled labor and capital, θ_3 is the elasticity of substitution across value-added and intermediate inputs, and θ_4 is the elasticity of substitution across intermediate inputs.

We consider a benchmark calibration with values for the elasticities $(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4) = (0.9, 0.67, 1.67, 0.5, 0.0001)$ informed by estimates from the microeconomic literature, which as we shall see, points to strong complementarities at the sectoral level ($\theta_4 = 0.0001$) and between value added and intermediates ($\theta_3 = 0.5$). For the elasticities of substitution between skilled labor and capital and between unskilled labor and the composite factor consisting of high-skilled labor and capital and skilled labor, we pick $\theta_1 = 0.67$ and $\theta_2 = 1.67$ which we are consistent with the microeconomic literature surveyed in reported by Johnson (1997) and Hamermesh (1996). We set the elasticity of substitution in consumption $\theta_0 = 0.9$, following Atalay (2017), Herrendorf et al. (2013), and Oberfield and Raval (2014), all of whom use an elasticity of substitution in consumption (across industries) of slightly less than one. For the elasticity of substitution across value-added and intermediate inputs, we set $\theta_3 = 0.5$. This accords with the estimates of Atalay (2017), who estimates this parameter to be between 0.4 and 0.8, as well as Boehm et al. (2019), who estimate this elasticity to be close to zero. Finally, we set the elasticity of substitution across intermediate inputs to be $\theta_4 = 0.001$, which matches the estimates of Atalay (2017). The aggregate production function arising endogenously from our model is different from the aggregate production function postulated by Krusell et al. (2000).

In the aggregate model of Krusell et al. (2000), macroeconomic elasticities of substitution between factors coincide with their microeconomic counterparts. As a result, they can be estimated using either microeconomic or macroeconomic data. Their particular choices come from a macroeconomic estimation with macroeconomic data using the first-order conditions of their aggregate model. They find that these estimates are consistent with microeconomic estimates in the microeconomic literature reported by Johnson (1997) and Hamermesh (1996).

By contrast, in our disaggregated model, the macroeconomic elasticities of substitution between factors are different from their microeconomic counterparts. We therefore rely entirely on microeconomic estimates to justify our choices the microeconomic elasticities of substitution θ_1 and θ_2. If one takes the view that the economy is described by our model, then their macroeconomic model is mis-specified. Our justification for choosing the same values as them for these elasticities is only that these values are consistent with estimates from the microeconomic literature, and not because they come out of the estimation of their possibly mis-specified macroeconomic model with macroeconomic data.

We use input-output tables and the integrated industry-level production account (KLEMS) data from the BEA. These two datasets report, at the industry level, the expenditures of each industry on different types of inputs (supplied from other industries), as well as the compensation of college (high-skill) employees and non-college (low-skill) employees. We attribute the remainder of each industry's value-added net of compensation of employees to the industry's capital stock.

Although we calibrate the model with industry-level data, the general methodology can accommodate as much data as available to the researcher. In particular, going from an industry-level model, to a firm or product level model is, conceptually, a very easy step given the generality of Propositions 2 and 7.

Production Function				Cost Function			
	Capital	Non-college	College		Capital	Non-college	College
Capital	–	1.67	0.67	Capital	–	1.67	0.67
Non-college	1.04	–	0.89	Non-college	1.26	–	1.09
College	0.67	1.67	–	College	0.67	1.67	–

Table 1: MESs between factors in the aggregate production function and in the aggregate cost function for the aggregate model.

Production Function				Cost Function			
	Capital	Non-college	College		Capital	Non-college	College
Capital	–	1.43	0.69	Capital	–	1.47	0.72
Non-college	0.94	–	0.94	Non-college	1.09	–	1.09
College	0.66	1.59	–	College	0.64	1.54	–

Table 2: MESs between factors in the aggregate production function and in the aggregate cost function for the disaggregated model.

In Table 1, we show the MES between the three primary factors for the aggregate model of Krusell et al. (2000) using both the production function (quantity elasticities) and the cost function (price elasticities). Since the aggregate model has a nested CES production function, the MES between capital and college labor in the aggregate production function and in the aggregate cost function are just equal to the corresponding micro elasticity $\theta_1 = 0.67$. Similarly, the MESs between either capital or college labor and non-college labor are just equal to the corresponding micro elasticity $\theta_2 = 1.67$.

In Table 2, we show the same macroeconomic elasticities of substitution (for both quantities and prices) using the disaggregated model. In the disaggregated model, MESs between factors depend not only on their microeconomic counterparts θ_1 and θ_2 as in the aggregate model, but also on the other microeconomic elasticities of substitution θ_0, θ_3, and θ_4. Note that the values of σ^C_{HK} and σ^F_{HK} end up being very similar to those used by Krusell et al. (2000) (respectively 0.64 vs. 0.67 and 0.66 vs. 0.67) : this is because some of the microeconomic elasticities are higher ($\theta_0 = 0.9$ and $\theta_2 = 1.67$) and some are lower ($\theta_3 = 0.5$ and $\theta_4 = 0.0001$), and their corresponding effects basically cancel out. The difference is bigger in the case of σ^C_{LK} and σ^F_{LK} (respectively 1.09 vs. 1.26 and 0.94 vs. 1.04) and the disaggregated model has lower values for these elasticities in part simply because in this case the benchmark microeconomic elasticity is higher to begin with.

To demonstrate this dependence in more detail, we show how the MES between capital and skilled labor changes in response to changes in the underlying microeconomic elasticities of substitution in Table 3, and how the MES between capital and skilled labor changes in response to changes in the underlying microeconomic elasticities of substitution in Table 4. In both tables, since the MESs in the aggregate cost function are weighted averages of the underlying micro-elasticities of substitution, the sum of the derivatives along this column add up to 1. The same is not true of the MESs in the aggregate production function.

	Aggregate Model		Disaggregated Model	
	Cost	Production	Cost	Production
σ_{HK}	0.67	0.67	0.64	0.66
Consumption: θ_0	–	–	0.13	0.12
VA vs. INT: θ_3	–	–	0.1	0.09
INT: θ_4	–	–	0.09	0.08
VA outer nest: θ_2	0	0	0.01	0.05
VA inner nest: θ_1	1	1	0.67	0.65

Table 3: Derivatives of the MES between high-skilled labor and capital w.r.t. micro elasticities of substitution.

We can use our estimates for the values of the macro elasticities of substitution to revisit the question posed by Krusell et al. (2000) and assess by how much the growth in capital due to the decline in the relative price of equipment has contributed to the widen-

	Aggregate Model		Disaggregated Model	
	Cost	Production	Cost	Production
σ_{LK}	1.28	1.05	1.09	0.93
Consumption: θ_0	–	–	0.11	0.14
VA vs. INT: θ_3	–	–	0.09	0.11
INT: θ_4	–	–	0.1	0.07
VA outer nest: θ_2	0.4	0.97	0.26	0.57
VA inner nest: θ_1	0.6	0.24	0.46	0.22

Table 4: Derivatives of the MES between low-skilled labor and capital w.r.t. micro elasticities of substitution.

ing skill premium. Following them, we ask how factor income shares would change if growth in capital had not accelerated in the mid-1970s. Absent this acceleration in the growth rate, the capital stock in 2015 would be lower. Specifically, maintaining the growth in the capital stock at its 1970s growth rate would be tantamount to a shock of size $d \log K = -0.37$ to today's capital stock.

In Table 5, we show the implied change in factor income shares in response to such a shock, for both the aggregate and disaggregated models. In the aggregate counterfactual economy, the capital share would be higher by about 5%, while the non-college share of income would be 7% higher, and the college share of income would be 13% lower. On the other hand, in the disaggregated counterfactual economy, the capital share would be 6% higher, while the non-college share of income would be 3% higher, and the college share would be 12% lower. Hence, viewed through the lens of the disaggregated economy, and compared to the aggregate economy, the accelerated growth in capital has weighed down the capital share more, increased the college share less, and weighed down the non-college share less (the last effect has been cut by more than half).

	Capital	Non-college	College
Aggregate model	0.05	0.07	-0.13
Disaggregated Model	0.06	0.03	-0.12

Table 5: The (log point) change in factor income shares in response to the shock $d \log K = -0.37$ in the aggregated and disaggregated model.

The change in the skill premium, defined as the change in the relative income of college and non-college labor, can also be estimated using the numbers in Table 1 and 2. Specifically, we can write

$$\left(\frac{1}{\sigma_{HK}} - \frac{1}{\sigma_{LK}}\right) \times \mathrm{d}\log K = \left(\frac{\mathrm{d}\log \Lambda_H}{\mathrm{d}\log K} - \frac{\mathrm{d}\log \Lambda_L}{\mathrm{d}\log K}\right) \times \mathrm{d}\log K,$$

putting the change in skill premium in the aggregate model at -0.20 log points and -0.16 log points in the disaggregated model.

The implication is that the disaggregated model reduces by 20% the increase in the skill premium that can be ascribed to the channel running through capital-skill complementarity and the decrease in the relative price of equipment identified by Krusell et al. (2000). This is of course by no means the final word on this question. The broad lesson is that disaggregating matters for some questions and less for others. There is a priori no way to know without doing the analysis. The point we wish to demonstrate is how easy it is to port prior analysis to more realistic production structures through a mechanical application of Propositions 1 and 6 — the only limit is the availability of data.