NBER WORKING PAPER SERIES

MICRO PROPAGATION AND MACRO AGGREGATION

David Baqaee
Elisa Rubbo

Working Paper 30538
http://www.nber.org/papers/w30538

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2022, Revised May 2023

We dedicate this paper to Emmanuel Farhi, whose work forms the basis of much the material we review and whose untimely passing precluded him from being an author of this review. We thank Ariel Burstein, Kilian Huber, Yasutaka Koike-Mori for comments. We thank Erin Eidschun and Gabriele Romano for excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Micro Propagation and Macro Aggregation
David Baqaee and Elisa Rubbo
NBER Working Paper No. 30538
October 2022, Revised May 2023
JEL No. E01

ABSTRACT

This paper reviews a framework for studying the aggregation and propagation of microeconomic shocks in general equilibrium. We discuss the determinants of aggregate measures of real economic activity, like real GDP, real domestic absorption, and aggregate productivity in both efficient and inefficient environments. We also discuss how shocks from one set of producers are transmitted to other producers through prices and quantities. The framework we provide is amenable to generalization and can be used to study any collection of producers ranging from one isolated producer, to an industry consisting of heterogeneous producers, to an entire economy. We conclude with a brief survey of some of the applied questions that can be addressed using the analytical tools presented in this review and avenues for future work.

David Baqaee
Department of Economics
University of California at Los Angeles
Bunche Hall
Los Angeles, CA 90095
and CEPR
and also NBER
baqaee@econ.ucla.edu

Elisa Rubbo
Booth School of Business
University of Chicago
5807 S. Woodlawn Avenue
Chicago, IL 60637
and NBER
elisarubbo@uchicago.edu

Editable slides for this paper are available on the authors' websites.
https://drive.google.com/file/d/1ctYIVdpL2oN3gKeIbfeo6XD-yGTx5BzH/

1 Introduction

Aggregation and propagation are two of the central problems in economics. The aggregation problem requires understanding how microeconomic disturbances affect aggregate variables. For example, how do shocks to one sector affect aggregate output, taking into account interactions between sectors? The propagation problem, on the other hand, requires understanding how shocks to one set of agents are transmitted, through prices and quantities, to other agents in the economy. For example, how do supply chain disruptions affect upstream suppliers and downstream consumers?

Historically, detailed and comprehensive data on individual firm and consumer behavior was rare or nonexistent. Therefore, much of the research on aggregation focused on conditions under which heterogeneity could either be ignored or boiled down to some low-dimensional subspace and disciplined with one or two aggregate moments. For the most part, this theoretical literature on aggregation yielded "negative" results. Except under very strong, and very counterfactual, assumptions, representative consumers and aggregate production functions do not exist. This left macroeconomics in an awkward position.

Theoretically, economies with heterogeneous consumers and firms do not need to qualitatively or quantitatively behave like those with a representative household and firm. Accounting for the disaggregated details of production substantively changes answers to many important questions. For example, the decomposition of aggregate productivity growth into technical and allocative efficiency, the economy's distance from the efficient frontier due to microeconomic distortions, the welfare gains from international trade, the slope of the Phillips curve and societal losses from price rigidity are all qualitatively and quantitatively affected by how well one models the production structure. But in practice, without disaggregated data to limit the range of possibilities, allowing for arbitrary forms of heterogeneity would result in theories with little predictive content where "anything goes."

In recent times, highly disaggregated microeconomic data have become available that can be used to discipline disaggregated models. New datasets allow us to trace, at very disaggregated levels and high frequencies, the transmission, propagation, and amplification of shocks. However, a flexible theoretical framework is necessary to make sense of such vast amounts of data, and to exploit its potential to not only revisit older questions but to also ask new questions. This article reviews and synthesizes some recent advances in this direction. We discuss a unified framework for studying the aggregation and propagation of shocks in disaggregated and inefficient environments, with a specific focus on shocks originating from the producers' side of the economy.[1] We also survey the wide-ranging applications of this framework, which go from international trade, to economic growth, to business cycle analysis.

In Section 2, we set up an abstract economic environment and define a data-consistent notion of aggregate economic activity. Our definition nests most aggregated real measures like real GDP, real consumption, real domestic absorption, industrial output, aggregate multifactor productivity, and so on. We then introduce the types of disturbances, or shocks, that we address in the paper. We also discuss the two types of questions that we focus on throughout the paper: first, what are the effects of changes in technologies and distortions on aggregate economic activity; second, how do changes in technology and distortions propagate from one group of producers to the rest. We show that the answers to these two seemingly different questions are intimately linked.

Most importantly, Section 2 presents a fundamental aggregation theorem, building on Hulten (1978), that applies to a much wider range of outcomes than real GDP. This result shows that around an efficient equilibrium, changes in the allocation of resources (appropriately defined) do not affect aggregate economic activity. Moreover, up to a first-order approximation, the aggregate impact of changes in external inputs and technologies are entirely determined by easily observable initial (pre-shock) sales shares. Conditional on these sales shares, aggregate outcomes do not depend on any other micro-level details. The rest of the paper focuses on departures from this benchmark result, and the difficulties of aggregation can be understood in terms of these departures.

To go further, in Section 3, we specialize the abstract environment of Section 2 to a fully spelled-out workhorse input-output model with a representative consumer and a single factor of production. Assuming a representative household and a single factor of production means that we set aside the interesting economics of aggregating across heterogeneous households and factors. We do this for expositional clarity, since the results we introduce can readily be generalized to environments with more factors and households. That is, we use this sandbox economy to provide concrete illustrations of how the fundamental aggregation theorem in Section 2 can be generalized but point interested readers to other sources for the more general treatment.

In Section 4, using the sandbox economy of Section 3, we characterize the microeco-

[1]Many shocks originating from the consumer or financial side of the economy can also fit in our framework. We do not illustrate this in detail, but we briefly point interested readers to the relevant sources.

nomic propagation of shocks. We derive propagation equations that show how shocks are transmitted through forward and backward linkages between sellers and buyers. The forward propagation equations describe how changes in costs are pushed forward from suppliers to their customers. The backward propagation equations determine how changes in demand are pushed backward from customers to their suppliers. These equations pin down the general equilibrium response of microeconomic variables like sales, quantities, and prices to primitives.

Moving from the forward and backward equations, Sections 5 and 6 connect the micro-level results on propagation to macro-level aggregation. Section 5 discusses aggregation in inefficient economies and Section 6 considers nonlinearities. In both cases, the response of aggregate output is tied to how certain sales shares respond to shocks, and the forward and backward linkage equations from Section 4 allow us to express changes in sales shares in terms of primitives. That is, the backward and forward propagation equations describe not just micro-level propagation but also macro-level aggregation.

Section 5 considers aggregation when the initial allocation is inefficient. In this case, reallocation effects are no longer irrelevant to a first-order. Any shock that reallocates resources towards producers that were too small from a social perspective increases aggregate productivity and output. We show that these reallocation effects can be tracked via changes in factor income shares, which depend on the micro-level propagation patterns pinned down by backward and forward equations. This means that input-output linkages directly matter, over and above the initial sales shares, for how resources are redirected in response to shocks.

In Section 6, we show that changes in sales shares not only inform aggregation in inefficient economies, but they also determine the nonlinear response of aggregate output to productivity shocks and distortions. Specifically, if sales rise for parts of the economy that are hit by negative productivity shocks, which happens in the case of complementarities, then aggregate output is concave with respect to such shocks. Therefore, once again, the forward and backward equations, which pin down changes in sales at the micro-level, are needed to understand the response of aggregate output. Similarly, the costs of a distorting wedge, in general equilibrium, are related to how much the quantity of the corresponding good responds. The more quantities respond, the larger the associated deadweight loss triangles, and the larger the reduction in output. Once again, the forward and backward equations can be used to express macro-level changes in terms of micro-level primitives.

In Section 7, we consider two extensions: allowing for endogenous labor supply and endogenous wedges using sticky prices. In each case, we discuss how the microeconomic details of the production structure matter, and how the aforementioned results can be used to study counterfactuals.

We postpone our discussion of the related literature until the end of the paper in Section 8. This allows us to explain how the basic framework we sketched out is related to different literatures, ranging from trade, to growth and development, to business cycles. We provide a bird's-eye view of the related literature and, for each topic, suggest further reading and avenues for future work.

2 General Framework and a Benchmark Result

In this section we define the macro- and micro-level questions addressed in the paper using a general and abstract economic environment, and we provide a benchmark aggregation result for efficient economies. In the rest of the paper, we discuss these questions and illustrate departures from our benchmark aggregation result using a fully specified model.

2.1 Environment

Consider some universe of economic actors consisting of producers, consumers, and factor endowments, potentially operating across different spatial regions and time periods. Let $\mathcal{U}$ denote the universe of goods produced and consumed by these actors. One of our central objective is studying how measures of aggregate real activity (such as GDP, GNE, the Solow residual...) for a subset of these actors change as a function of micro-level primitives. For now, we do not explicitly enumerate the relevant primitives. Primitives could include technologies, distorting wedge-like taxes, shocks to consumer preferences, factor endowments, and so on. We index all these primitives, which we have not fully spelled out, by some scalar s and refer to s as the *state* of the economy.

Below, we introduce the assumptions necessary to define aggregate activity for a "sub-economy" of interest (Assumption 1 and Assumption 2), without explicitly spelling out all the details required to pin down the equilibrium. Our definition includes most common measures of aggregate activity. As we will see, these measures differ in their definitions of the sub-economy of interest, its final outputs, and its external inputs. First, we assume there are market prices for each good, though we do not spell out how these prices are determined.

Assumption 1 (Existence of Markets)**.** *For every state s, there are prices $p(s) \in \mathbb{R}^{|\mathcal{U}|}$.*

Next, we consider a subset of goods $\mathcal{N} \subset \mathcal{U}$, and for each $i \in \mathcal{N}$, we define a quantity y_i to be called *final output*. We also define a set of *external inputs* $\mathcal{H} \subset \mathcal{U}$, and denote the total quantity of external input $k \in \mathcal{H}$ by l_k. We assume that production of final outputs can be represented in the following way.

Assumption 2 (Production and Resource Constraints)**.** *For each good $i \in \mathcal{N}$, there is a function f_i such that*

$$x_i = f_i\left(\{x_{ij}\}_{j\in\mathcal{N}}, \{l_{ij}\}_{j\in\mathcal{H}}; A_i\right), \tag{1}$$

where x_i is total output of good i, while x_{ij} and l_{ik} are the quantities of good j and external input k used to produce i. Moreover, for every $i \in \mathcal{N}$ and $k \in \mathcal{H}$, the following equations hold

$$x_i = y_i + \sum_{j\in\mathcal{N}} x_{ji}, \quad \textit{and} \quad \sum_{i\in\mathcal{N}} l_{ik} = l_k.$$

Finally, for every $i \in \mathcal{N}$, the input combinations in (1) *are chosen to minimize costs given prices.*

That is, the total output x_i of good i equals the sum of the final uses we wish to aggregate, y_i, as well as intermediate inputs use by the collective, $\sum_{j\in\mathcal{N}} x_{ji}$. We refer to l_k as the supply of external input k. Any shifters in f_i, here labelled by A_i, are called *technology*.

Nominal aggregate activity is just the total nominal value of final outputs.

Definition 1 (Nominal Aggregate Activity)**.** As a function of the state s, *nominal economic activity* in $\mathcal{N}$ is given by total nominal final output

$$E(s) = \sum_{i\in\mathcal{N}} p_i(s) y_i(s).$$

The change in *real economic activity* is the change in final outputs minus external inputs measured at constant prices.[2]

Definition 2 (Real Aggregate Activity)**.** As the state changes, the infinitesimal change in *real economic activity* is

$$\frac{d\log Y}{ds} = \sum_{i\in\mathcal{N}} \frac{p_i(s)}{E(s)} \frac{dy_i}{ds} - \sum_{i\in\mathcal{H}} \frac{p_i(s)}{E(s)} \frac{dl_i}{ds}. \tag{2}$$

[2]This is how quantity indices are defined and measured in the data. Equation (3) is the continuous-time analog to chain-weighted quantity indexes used in the national accounts.

The non-infinitesimal change in real activity between two states s_0 and s_1 is defined by

$$\Delta \log Y = \log Y(s_1) - \log Y(s_0) = \int_{s_0}^{s_1} \frac{d \log Y}{ds} ds. \tag{3}$$

By altering how y_i is mapped to data, E can measure many different notions of nominal activity, which in turn give rise to multiple notions of real activity $d \log Y$, depending on the set of external inputs $\mathcal{H}$. For real quantity indices, like real GDP, real GNE (gross national expenditure or domestic absorption), real consumption, or real investment, the set of external inputs is empty. For productivity-type measures, instead, like the Solow residual, the set of external inputs is non-empty. This is because productivity-type measures subtract some notion of input growth from output growth. For example, the Solow residual subtracts capital and labor growth from real GDP growth. Gross industrial productivity subtracts capital, labor, and materials growth from gross industrial output growth. Value-added productivity growth subtracts labor and capital growth from real value-added growth.

Many standard measures of real economic activity can be represented in this way. We provide some specific examples below.

Example 1 (GDP)**.** For gross domestic product (GDP), the set of goods is $\mathcal{N} = \mathcal{U}$, and the external input set is empty $\mathcal{H} = \emptyset$. Final outputs y_i are given by consumption, investment, government spending, plus net exports of each good i in a given country and year. That is, for every $i \in \mathcal{N}$ define

$$y_i = c_i + g_i + I_i + x_i^X - x_i^M,$$

where c_i, g_i, I_i, x_i^X, and x_i^M are consumption, government spending, investment, exports and imports of good i. Note that it is possible for y_i to be negative if total imports exceed $c_i + g_i + I_i + x_i^X$, which can happen if i is used as an intermediate input. Next, let f_i denote the production function of good i for the domestic economy:

$$x_i = f_i\left(\{x_{ij}\}_{j \in \mathcal{N}}, A_i\right).$$

Two remarks are in order. First, if the domestic country is incapable of producing good i, then f_i is just a zero function. Second, domestic primary factors, such as labor, land, and the initial stock of capital, are also included in the set $\mathcal{N} = \mathcal{U}$, with $y_i = 0$. Their production functions have zero returns to scale, $f_i = A_i$, and here, A_i is a technology parameter which controls the supply or productivity of that primary factor. Note that Assumption 2 is satisfied, because

$$x_i = y_i + \sum_{j \in \mathcal{N}} x_{ji},$$

for every i. Moreover, in the language of Assumption 2, the technology parameters A_i include the supply of domestic primary factors.

Example 2 (Solow Residual)**.** To capture the Solow residual, we use the same definitions as for real GDP, but we now define the set of external inputs $\mathcal{H}$ to be domestic primary factors. Typically, this is domestic labor and capital.

Example 3 (GNE)**.** Unlike GDP and the Solow residual, which are measures of production and productivity, GNE (also called domestic absorption) is a measure of consumption. To capture GNE, define

$$y_i = c_i + g_i + I_i,$$

where c_i, g_i, I_i are consumption, government, and investment uses of good i in the domestic country in a given year. Again, denote by x_i^M and x_i^X imports and exports of good i. As an accounting identity, the domestic absorption of good i is equal to domestic production plus imports minus exports and intermediate input use:

$$y_i = Z_i g_i \left(\{x_{ij}\}_{j \in \mathcal{N}} \right) + x_i^M - \sum_{j \in \mathcal{N}} x_{ji} - x_i^X.$$

Here, g_i is the domestic production function for good i, which may be zero if the domestic economy does not produce good i. Intermediate inputs used to produced i are x_{ij} and Z_i is a parameter controlling the domestic production technology. Again, we can put this into a formulation where Assumption 2 holds as follows:

$$x_i = y_i + \sum_{j \in \mathcal{N}} x_{ji} = Z_i g_i \left(\{x_{ij}\}_{j \in \mathcal{N}} \right) + x_i^M - x_i^X = f_i \left(\{x_{ij}\}_{j \in \mathcal{N}}, A_i \right)$$

Here, $A_i = (Z_i, x_i^M, x_i^X)$ are parameters of the function f_i, which means that imports and exports are part of the "technology" parameters in the terminology of Assumption 2.

Example 4 (Net present value of consumption)**.** Consider a typical neoclassical growth model with a single consumption good in each period. To define the net present value

(NPV) of consumption, let the set of external inputs be $\mathcal{H} = \emptyset$, and let $\mathcal{N} = \mathcal{U}$ be the set of consumption goods and labor in each period. We distinguish between the consumption good and the labor good in each period using superscripts C and L.

The prices p_i^C and p_i^L correspond to the Arrow-Debreu price of consumption and labor in each period. Therefore, Assumption 1 is satisfied. For real economic activity $d \log Y$ to capture changes in the net present value of consumption or welfare, let y_i^C and y_i^L denote consumption in period i of the good and of labor. Since the household does not directly consume labor, $y_i^L = 0$ for every i. We now show how to define production functions in such a way that Assumption 2 is satisfied.

The production function for the labor good x_i^L is just an exogenous zero returns to scale function with some shifter A_i^L controlling the endowment of labor for each period. For the consumption good in period i, the resource constraint is

$$y_i^C + x_{i+1,i}^C = x_i^C,$$

where x_i^C is total non-labor wealth, y_i^C is consumption, and $x_{i+1,i}^C$ is the capital stock in period $i+1$. The production function satisfies

$$x_i^C = f_i\left(\left\{x_{ij}^C, x_{ij}^L\right\}_{j\in\mathcal{N}}; A_i\right) = A_i^C g_i(x_{i,i-1}^C, x_i^L) + (1-\delta)x_{i,i-1}^C,$$

where δ is the depreciation rate of the capital stock. In the language of Assumption 2, the parameters A_i^C and δ are technology shifters, and the capital stock is treated as an intermediate input.

Other examples of activity measures that satisfy our definition are personal consumption expenditures (PCE), where y_i corresponds to consumption of both domestic and imported goods in a given year, and aggregate investment, where y_i corresponds to investment goods purchased by domestic producers in a given year.

Distortions. The allocations in our economy need not be efficient. We define wedges in our sub-economy of interest, $\mathcal{N}$, as follows.

Definition 3 (Wedges)**.** For each good $i \in \mathcal{N}$, the *wedge* μ_i is defined to be the price, p_i, divided by the marginal cost of producing i.[3]

[3] We do not define wedges for i in the special case where f_i has zero returns to scale because in this case marginal cost is undefined. Having said that, our notation for wedges is general and allows us to

As mentioned earlier, Assumptions 1 and 2 and Definition 3 are not enough information to pin down equilibrium allocations. For example, we have not specified how wedges and prices are determined. Furthermore, we have not imposed that y_i necessarily be a positive number — given the limited assumptions we impose, y_i may be positive or negative. For example, as we saw in Example 1, when E is nominal GDP, then y_i is negative whenever i is an imported intermediate good.

Table 1 illustrates how most measures of real economic activity, including real GDP, real GNE, industrial output, and the Solow residual, can be captured by Definition 2 in a way that Assumptions 1 and 2 are satisfied. The distinction between different measures comes from their definition of final outputs, y, and external inputs, l, which in turn determines the sets $\mathcal{N}$ and $\mathcal{H}$. As we saw above, in each case, the functions $\{f_i\}_{i\in\mathcal{N}}$ need to be defined appropriately.

We now discuss the two types of questions we study in the rest of the paper: macro-level aggregation and micro-level propagation, starting with aggregation.

2.2 Aggregation

A first set of questions concerns quantifying the microeconomic drivers of aggregate growth. The growth accounting literature, pioneered by Solow (1957), Domar (1961) and Hulten (1978), focused on perfectly competitive economies. Hall (1990), Basu and Fernald (1997), and Levinsohn and Petrin (2003) propose different approaches to generalizing growth accounting beyond marginal-cost-pricing environments.

In this paper we build on the approach in Baqaee and Farhi (2020c). We define aggregate growth as the change in our chosen measure of real activity, and decompose it into three sources: (1) increases in the supply of external inputs to $\mathcal{N}$, (2) technological progress within $\mathcal{N}$, and (3) changes in the allocation of resources across producers in $\mathcal{N}$.

To formalize our decomposition, we introduce the *allocation rule* $\mathcal{X}$, which dictates, in a way we will make precise later in Section 5, the share of each resource going to each use of that resource in $\mathcal{N}$. Technology parameters A, external inputs l, and the allocation rule, $\mathcal{X}$, jointly pin down the quantity of all final outputs and external inputs. Hence, we

fully saturate the model with wedges to reach any desired allocation. For example, to capture buyer-seller specific wedges, we simply introduce a fictitious middle-man between buyers and sellers. A markup charged by this fictitious middleman is isomorphic to a buyer-seller specific wedge. Wedges on endowment goods are irrelevant since they do not distort allocations.

Table 1: Alternative Definitions of Real Economic Activity

$d \log Y$	Set of quantities y_i	Set of actors $\mathcal{N}$	External inputs $\mathcal{H}$
Real GDP	Domestic consumption + investment + government spending + net exports of all goods $i \in \mathcal{U}$	Domestic producers and primary factors in a given year	Empty set
Aggregate multifactor productivity (Solow residual)	Domestic consumption + investment + government spending + net exports of all goods $i \in \mathcal{U}$	Domestic producers in a given year	Labor and capital inputs
Real Gross National Expenditures (Domestic Absorption)	Domestic consumption + investment + government spending of all goods $i \in \mathcal{U}$	Domestic consumption, investment, and government goods	Empty set
Personal Consumption Expenditures (PCE)	Consumption of domestic and imported goods $i \in \mathcal{U}$ in a region-year	Producers and net imports of consumption goods in a period	Empty set
Net present value of real consumption	Consumption of all goods $i \in \mathcal{U}$ in a region across periods	Producers of consumption goods across multiple time periods	Empty set
Aggregate investment	Investment goods purchased by domestic producers in a given yea	Domestic investment goods	Empty set
Real industry output	Products of the industry	Producers and factors in the industry-year	Empty set
Real industry value added	Products of the industry minus intermediates	Producers and factors in the industry-year	Empty set
Gross industrial productivity	Products of the industry	Producers in the industry-year	Labor and capital and intermediate inputs
Value-added industrial productivity	Products of the industry minus intermediates	Producers in the industry-year	Labor and capital inputs

can decompose the change in real activity Y, to a first-order, into three parts:

$$\frac{d\log Y}{ds} = \underbrace{\frac{\partial \log Y}{\partial \log l}\frac{d\log l}{ds}}_{\Delta \text{ external inputs}} + \underbrace{\frac{\partial \log Y}{\partial \log A}\frac{d\log A}{ds}}_{\Delta \text{ technology}} + \underbrace{\frac{\partial \log Y}{\partial \mathcal{X}}\frac{d\mathcal{X}}{ds}}_{\Delta \text{ allocation}}, \tag{4}$$

where the partial derivative in each summand holds fixed the other two effects.[4] So, for example, the first summand considers changes in the quantity of external inputs, holding fixed internal technology parameters A and the allocation rule $\mathcal{X}$.

The Fundamental Theorem of Aggregation. Theorem 1 is a fundamental aggregation result that serves as a benchmark for much of the rest of the paper. This result, which is a generalization of Hulten (1978), decomposes aggregate productivity along the lines of (4). The pay-off to our abstract approach is that Theorem 1 simultaneously applies to all notions of aggregate real activity outlined in Table 1, including measures of production, consumption, and productivity. It also applies to both intertemporal and open economies.

Theorem 1 (Aggregation in Efficient Environment)**.** *In the absence of wedges in $\mathcal{N}$, that is if $\mu_i = 1$ for every i, we have*

$$\frac{d\log Y}{ds} = \underbrace{0}_{\Delta \text{ external inputs}} + \underbrace{\sum_{i\in\mathcal{N}} \frac{p_i x_i}{E}\frac{\partial \log f_i}{\partial \log A}\frac{d\log A}{ds}}_{\Delta \text{ technology}} + \underbrace{0}_{\Delta \text{ allocation}}. \tag{5}$$

In the equation above, $\partial \log f_i / \partial \log A$ is the mechanical increase in the production of i, holding fixed all inputs. If A is a Hicks-neutral productivity shifter for i, then $\partial \log f_i / \partial \log A = 1$.

Theorem 1 implies that, around an equilibrium with no wedges and to a first-order approximation, changes in aggregate real activity *only* depend on changes in technology. Changes in any other component of the state s (wedges, external inputs, or any supply and demand shocks outside of $\mathcal{N}$) have no direct effect on output.

We provide some explicit examples below, but first, we describe each term in (5) in sequence starting with external inputs. Intuitively, changes in the quantity of external

[4] In other words, for each of $Z \in \{l, A, \mathcal{X}\}$, using (2), we define

$$\frac{\partial \log Y}{\partial Z} = \sum_{i\in\mathcal{N}} \frac{p_i}{E}\frac{\partial y_i}{\partial Z} - \sum_{i\in\mathcal{H}} \frac{p_i}{E}\frac{\partial l_i}{\partial Z}.$$

inputs do not affect aggregate real activity because in (2) their value is subtracted out of aggregate output at cost. Since, for each input, the marginal revenue product is equal to the cost, the change in total output of the firms in $\mathcal{N}$ equals the change in inputs, with opposite signs. We provide a concrete illustration in Example 6 below.

Next, consider the effect of technology shocks inside $\mathcal{N}$. The contribution of technology shocks to each producer is proportional to that producer's sales as a share of nominal output, oftentimes called its *Domar weight*. Crucially, the Domar weight i is related to its total sales $p_i x_i$ rather than its final sales $p_i y_i$. When there are intermediate inputs, then total sales exceed final sales. This shows that intermediate inputs amplify the effect of technology shocks on real activity. Furthermore, shocks to goods whose total sales are zero, $p_i x_i = 0$, like imported intermediates, where $p_i y_i = -\sum_j p_i x_{ji}$, or goods that are excluded from $\mathcal{N}$, do not affect real economic activity. We provide a concrete illustration in Example 5 below.[5]

Finally, equation (5) shows that changes in how resources are allocated across different producers in $\mathcal{N}$ do not matter. This is because the marginal revenue product of each input is being equated across competing uses at the initial allocation. Therefore, a reallocation from one producer to another raises and lowers aggregate output by exactly off-setting amounts. In other words, reallocation effects are irrelevant regardless of their cause, be they technology shocks, demand shocks, or changes in wedges.

Theorem 1 nests many results in the literature. For example, Shephard's lemma, Hotelling's lemma, Hulten's theorem, the sufficient statistics formula used by Bachmann et al. (2022), and the open-economy version of Hulten's theorem in Burstein and Cravino (2015) and Baqaee and Farhi (2019a) can all be derived as special cases. Theorem 1 also applies to other measures of economic activity, like real GNE, which is a measure of welfare in an open economy. Theorem 1 generalizes these results since it allows the real activity measure to be the output of any subset of producers, and the state s can index any variable, determined either inside or outside the set $\mathcal{N}$.

To better understand Theorem 1, it helps to think through the following examples from Table 1.

Example 5 (Determinants of Real GDP)**.** Consider an efficient economy with a single consumption good produced using labor and an imported intermediate with constant returns to scale. We apply Theorem 1 to real GDP. The set of goods is $\mathcal{N} = \mathcal{U} = \{\text{cons, lab, nx}\}$,

[5] See Kehoe and Ruhl (2008), Burstein and Cravino (2015), and Baqaee and Farhi (2019a) for more information about how imports affect real GDP. As shown by Baqaee and Farhi (2019a), the irrelevance of imports to real GDP depends on productive efficiency, and breaks down in inefficient economies.

corresponding to consumption, labor, and net exports. In the definition of real GDP, there are no external inputs, so $\mathcal{H} = \varnothing$ (see Table 1).

The production function for the consumption good is

$$x_{cons} = f\left(Zx_{cons,lab}, x_{cons,nx}\right),$$

where Z is labor-augmenting technical change. The production function for labor has zero returns to scale:

$$x_{lab} = L,$$

where L is the endowment of labor. Finally, since domestic producers do not produce the imported good, the total production of nx by domestic producers is $x_{nx} = 0$. Therefore, the vector $A = (Z, L)$ captures the state of "technology" in the terminology of Assumption 2.

Following the notation above, the final goods vector is

$$y = \begin{pmatrix} y_{cons} \\ y_{lab} \\ y_{nx} \end{pmatrix} = \begin{pmatrix} x_{cons} \\ x_{lab} - x_{cons,lab} \\ x_{nx} - x_{cons,nx} \end{pmatrix} = \begin{pmatrix} x_{cons} \\ 0 \\ -x_{cons,nx} \end{pmatrix}.$$

Following Definition 1, nominal GDP is just consumption minus imports:

$$E = \sum_i p_i y_i = p_{cons} x_{cons} - p_{nx} x_{cons,nx}.$$

Following Definition 2, the change in real GDP is

$$d \log Y_{GDP} = \frac{p_{cons}}{E} \cdot dy_{cons} + \frac{p_{lab}}{E} \cdot dy_{\text{labor}} + \frac{p_{nx}}{E} dy_{nx}.$$

Theorem 1 implies that we can write

$$d \log Y_{GDP} = \frac{p_{lab} x_{lab}}{E} d \log L + \frac{p_{cons} x_{cons}}{E} \frac{\partial \log f}{\partial \log Z} d \log Z = d \log Z + d \log L,$$

where we use $p_{lab} x_{lab} / E = 1$. Hence, real GDP increases if, and only if, domestic technology or employment increases. On the other hand, and perhaps counterintuively, real GDP does not respond to shocks to imported intermediates, even if those inputs are essential for production.

Example 6 (Determinants of Solow Residual)**.** Consider the same economy as in the previous example, but now focus on the Solow residual instead of GDP. In this case, we let $\mathcal{N} = \mathcal{U} = \{\text{cons, nx}\}$. For the Solow residual, labor is treated as an external input. Hence, $\mathcal{H} = \{\text{lab}\}$, with an exogenous resource constraint given by $l_{\text{labor}} = L$. Following our notation, changes in the Solow residual net out external inputs, and are given by

$$d\log Y_{SR} = \frac{p_{cons}}{E} \cdot dy_{cons} + \frac{p_{nx}}{E} dy_{nx} - \frac{p_{lab}}{E} \cdot dl_{lab} = d\log Z,$$

which is a measure of domestic productive capacity purged of changes in employment. As per Theorem 1, external inputs have no effect on $d\log Y_{SR}$.

Example 7 (Determinants of Real GNE)**.** Consider real GNE, described in Example 3. For simplicity, assume there is no government or investment, so that domestic absoportion is simply domestic consumption. In this case, E is total nominal consumption by domestic consumers, and y_i is consumption of good i. Use the same notation as in Example 3, and denote domestic production of good i by $z_i = x_i + x_i^X - x_i^M$. This is equal to zero if the domestic country does not produce good i.

Applying Theorem 1 yields the following decomposition of real domestic absorption

$$\frac{d\log Y_{RGNE}}{ds} = \sum_{i \in \mathcal{N}} \frac{p_i x_i}{E} \left[\frac{p_i z_i}{p_i x_i} \frac{d\log Z_i}{ds} + \frac{p_i x_i^M}{p_i x_i} \frac{d\log x_i^M}{ds} - \frac{p_i x_i^X}{p_i x_i} \frac{d\log x_i^X}{ds} \right].$$

That is, changes in real GNE can be boiled down to sales-weighted changes in domestic technology Z plus imports x^M minus exports x^X. Furthermore, given (Z, x^M, x^X), which are "technologies" in the language of Assumption 2, other shocks and reallocations have no effect on domestic absorption.[6]

Example 8 (Determinants of Net Present Value of Consumption)**.** Consider again the neoclassical growth model, described in Example 4. Suppose that the state s controls the productivity shifters A_i^C in each period. According to Theorem 1, changes in the net present value of consumption, or welfare, are

$$\frac{d\log Y_{NPV}}{ds} = \sum_{i=1}^{\infty} \frac{p_i^C x_i^C}{E} \frac{\partial \log f_i}{\partial \log A_i^C} \frac{d\log A_i}{ds} = \sum_{i=1}^{\infty} \frac{Q_i GDP_i}{E} \frac{d\log A_i^C}{ds},$$

[6] Bachmann et al. (2022) use this logic, combined with methods in Section 6, to study the nonlinear impact of a Russian gas embargo on German real GNE.

where Q_i is the stochastic discount factor for period i and E is permanent income. That is, the elasticity of welfare to a productivity shock in period i is just the discounted GDP of period i relative to permanent income.[7] This fully accounts for the way that a productivity shock induces additional capital accumulation.

These examples demonstrate the powerful generality of Theorem 1 to different contexts. Theorem 1 does have some important limitations however. First, if prices are not always equal to marginal cost, the theorem does not predict output responses correctly. Second, there can be strong nonlinearities that make the first-order approximation in Theorem 1 unreliable even under the assumption of efficiency. Nevertheless, Theorem 1 provides a remarkably general point of departure for understanding the importance of distortions and nonlinearities. We discuss distortions and nonlinearities briefly below.

Distortions. Much of the simplicity of Theorem 1 derives from its assumption of productive efficiency. When there are distortions inside $\mathcal{N}$, modelled as gaps between prices and marginal costs, then changes in both external inputs and the allocation rule can have first order effects on aggregate output. First, changes in the total quantity of external inputs may affect output because the marginal revenue product of external inputs may not equal their cost. Second, even holding external inputs and technologies constant, changes in the allocation rule may affect output because a reshuffling of resources can raise or lower aggregate output. We consider these scenarios in Section 5.

Nonlinearities. The decomposition in Theorem 1 is linear. In the presence of strong nonlinearities, or large shocks, a first-order approximation can be misleading. One indicator of nonlinearity is asymmetry — for example, negative shocks to critical industries like oil or energy are much more damaging than positive shocks are beneficial due to strong complementarities. A linear expansion like (4) does not take this into account. Similarly, interactions between shocks are ruled out by first-order approximations. For example, a negative shock to electricity generation is likely to be much more important for aggregate output if it coincides with a negative shock to oil extraction. When relying on linear approximations, the effect of two simultaneous shocks is just the sum of the effect of each shock on its own. Understanding these nonlinear interactions is precisely one

[7] Basu et al. (2022) deploy a similar argument to show that the welfare of a country's infinitely-lived representative consumer is summarized, to a first order, by the discounted stream of Solow residuals and the initial capital stock.

of the reasons we want to build more disaggregated models of the aggregate economy. Section 6 considers nonlinear aggregation.

2.3 Propagation of Shocks

A second set of questions concerns how a change in the state s affects microeconomic prices and quantities in our sub-economy $\mathcal{N}$. The outcomes of interest could be the prices and quantities of specific final goods in y, individual external inputs in l, and intermediate inputs produced and used by firms in $\mathcal{N}$. Understanding these outcomes is useful for a wide variety of applications, such as the distributional impact of a shock on consumption or income, the comovement of variables with one-another, the propagation of sectoral shocks along supply chains in both prices and quantities, the response of factor income shares to changes in technology and market structure, and structural transformation.

Theorem 1 shows that in an efficient equilibrium and to a first-order approximation, propagation questions are unrelated to aggregation questions. We can use Theorem 1 without needing to know how prices and quantities of individual producers change in response to shocks. In the rest of the paper, we study aggregation and propagation in inefficient and nonlinear environments. As we shall see, in the presence of distortions and nonlinearities, the uncoupling between aggregation and propagation disappears.

In the next section, we introduce a simple model, and using this model, we discuss propagation, in Section 4, and aggregation, in Sections 5 and 6.

3 A Sandbox Economy

In this section, we introduce a toy or "sandbox" economy that we use for the rest of the analysis. The sandbox economy is less general than the environment in Section 2: it is closed, and it features a representative household with homothetic preferences, constant-returns production functions, and a single primary factor of production. We use this sandbox economy to illustrate more abstract results using a fully-spelled out model. We discuss in Section 8 how the ideas from the sandbox economy can be extended to economies with multiple households, multiple factors, and non-constant returns to scale. Despite these restrictions, the sandbox economy is still rich enough that it nests many standard models with intermediate inputs.

3.1 Primitives

There is a population of consumers with identical homothetic preferences and N producers. Household preferences, in money-metric terms, are given by some constant returns-to-scale function

$$C = \mathcal{C}(y_1, \ldots, y_N).$$

The ideal price index for the representative consumer is denoted by P.[8] We refer to goods produced with zero returns to scale (i.e. endowment goods) as *factors*. There is a single primary factor called labor and its supply is exogenous.[9] The budget constraint equates total final expenditures to labor and profit income,

$$E = \sum_i p_i y_i = wL + \sum_i \pi_i,$$

where w is the price of labor (wage), and π denotes the vector of profits. The production function of each $i \in N$ uses labor and intermediates to make each good i, as in (1)

$$x_i = A_i f_i \left(L_i, \{x_{ij}\}_{j \in N} \right), \tag{6}$$

where A_i is a Hicks-neutral productivity shock.[10] All production functions, except the one for labor, have constant returns to scale.

We focus on two popular notions of aggregate output: real GDP and the Solow residual (see Table 1). For this sandbox economy, real GDP is of particular interest because changes in real GDP exactly coincide with the change in consumer welfare in money-metric terms:

$$\Delta \log Y_{RGDP} = \Delta \log C.$$

Therefore, understanding real GDP is tantamount to understanding welfare. Furthermore, since the economy is closed and there is no investment, real GDP is also the same as domestic absorption and consumption.

In the language of Section 2, our universe of goods is $\mathcal{U} = \{1, \ldots, N\} \cup \{C\} \cup \{L\}$, where, with some abuse of notation, C denotes the final consumer and L denotes the

[8]The ideal price index is defined as $\min_{\boldsymbol{y}}\{\sum_i p_i y_i : C = 1\}$.

[9]Section 7.1 extends this setup to account for elastic labor supply.

[10]Input-biased technical change can be captured as a special case of this using fictitious intermediaries. To capture a technology shock to x_{ij}, we simply introduce a fictitious firm that buys from j and sells to i and a Hicks-neutral shock to this fictitious producer is isomorphic to an input-specific shock to x_{ij}.

primary factor. The collection of producers we are interested in is either $\mathcal{N} = \{1, \ldots, N\} \cup \{L\}$ or $\mathcal{N} = \{1, \ldots, N\}$ depending on whether we are interested in real GDP or the Solow residual.

3.2 Input-output Notation

We now introduce some useful notation. To write compact formulas, we treat final consumption and labor as additional producers. The final consumption sector buys the various goods and assembles them into the consumption bundle, which it then sells to the representative agent with $\mu_C = 1$. The labor sector is an endowment of labor sold to other producers with $\mu_L = 1$. We order $\mathcal{U}$ by putting the consumption sector first and the labor sector last.

The input-output matrix is the $(N+2) \times (N+2)$ matrix whose ij-th element is

$$\Omega_{ij} = \frac{p_j x_{ij}}{p_i x_i}.$$

In words, Ω_{ij} is i's expenditures on j as a share of i's sales. In a similar way, we define the *cost-based* input-output matrix $\tilde{\Omega}_{ij} = \mu_i \Omega_{ij}$. With our notation, the sales- and cost-based input-output matrices are as follows:

$$\Omega = \left[\begin{array}{c|ccc|c} 0 & \Omega_{C1} & \cdots & \Omega_{CN} & 0 \\ \hline 0 & \Omega_{11} & \cdots & \Omega_{1N} & \Omega_{1L} \\ 0 & & \ddots & & \\ 0 & \Omega_{N1} & & \Omega_{NN} & \Omega_{NL} \\ \hline 0 & 0 & \cdots & 0 & 0 \end{array}\right], \text{ and } \tilde{\Omega} = \left[\begin{array}{c|ccc|c} 0 & \Omega_{C1} & \cdots & \Omega_{CN} & 0 \\ \hline 0 & \mu_1\Omega_{11} & \cdots & \mu_1\Omega_{1N} & \mu_1\Omega_{1L} \\ 0 & & \ddots & & \\ 0 & \mu_N\Omega_{N1} & & \mu_N\Omega_{NN} & \mu_N\Omega_{NL} \\ \hline 0 & 0 & \cdots & 0 & 0 \end{array}\right].$$

We also define the sales-based and cost-based *Leontief inverses* as $\Psi = (I - \Omega)^{-1}$, and $\tilde{\Psi} = (I - \tilde{\Omega})^{-1}$. While the elements $\tilde{\Omega}_{ij}$ record the direct exposure of producer i's cost to j's prices, the Leontief inverse $\tilde{\Psi}$ records the direct and indirect exposures through the production network. This can be seen most clearly by writing

$$\tilde{\Psi} = \left(I - \tilde{\Omega}\right)^{-1} = I + \tilde{\Omega} + \tilde{\Omega}^2 + \ldots,$$

and noting that $(\tilde{\Omega}^n)_{ij}$ measures the weighted sums of all paths of length n from producer i to producer j.

Finally, denote the total sales of i relative to total nominal output, or the Domar weight, by

$$\lambda_i = \frac{p_i x_i}{\sum_{j\in\mathcal{N}} p_j y_j}.$$

Labor plays an important role in our analysis, and hence we use a special notation for the Domar weight of labor denoting it by $\Lambda \equiv \lambda_L$.

With our notation, the sales- and cost-based Leontief inverses are as follows:

$$\Psi = \left[\begin{array}{c|ccc|c} 1 & \lambda_1 & \cdots & \lambda_N & \Lambda \\ \hline 0 & \Psi_{11} & \cdots & \Psi_{1N} & \Psi_{1L} \\ 0 & & \ddots & & \\ 0 & \Psi_{N1} & & \Psi_{NN} & \Psi_{NL} \\ \hline 0 & 0 & \cdots & 0 & 1 \end{array}\right], \text{ and } \tilde{\Psi} = \left[\begin{array}{c|ccc|c} 1 & \tilde{\lambda}_1 & \cdots & \tilde{\lambda}_N & 1 \\ \hline 0 & \tilde{\Psi}_{11} & \cdots & \tilde{\Psi}_{1N} & 1 \\ 0 & & \ddots & & \\ 0 & \tilde{\Psi}_{N1} & & \tilde{\Psi}_{NN} & 1 \\ \hline 0 & 0 & \cdots & 0 & 1 \end{array}\right].$$

The first row of the Leontief inverse Ψ, corresponding to final consumption, contains the Domar weights. To see why this is the case, note that the aggregate budget constraint implies $\lambda_C = 1$, while the resource constraints for each $i \in \{1, \ldots, N\} \cup L$ imply that total sales of i equal total purchases from producers and final consumers:

$$\lambda_i = \Omega_{Ci} + \sum_{j=1}^{N} \lambda_j \Omega_{ji} = \sum_{j\in\mathcal{U}} \lambda_j \Omega_{ji}. \tag{7}$$

Equation (7) in turn implies $\lambda_i = \Psi_{Ci}$.[11] In particular, the resource constraint for labor implies that

$$\Lambda = \frac{wL}{\sum_j p_j y_j} = \sum_{j\in\mathcal{U}} \lambda_j \Omega_{jL} = \Psi_{CL}.$$

The last column of Ψ captures the fraction of each producer's revenues that eventually is paid to labor. If there were no markups, then we would have $\Psi_{iL} = \tilde{\Psi}_{iL} = 1$ for every i. If there are positive markups ($\mu > 1$), then $\Psi_{iL} < 1$, because part of revenues are paid off as profits rather than as wages at each step of the supply chain. One can thus interpret Ψ_{iL} as a measure of the total inverse markup along i's supply chain (including i itself). In particular, the equality $\Psi_{CL} = \Lambda$ highlights that the labor share can be interpreted as a measure of the aggregate inverse markup. The last column of $\tilde{\Psi}$ instead captures the

[11] We can rewrite (7) as $\lambda_i = \sum_{j\in N} \Omega_{Cj} \Psi_{ji} = \Psi_{Ci}$, where the last equality follows from the fact that $\Omega_{iC} = 0$ for every i.

total fraction of each producer's costs that accrues to labor, taking into account indirect payments through the supply chain. Since labor is the only primary factor, all costs ultimately originate from wage payments, either directly or indirectly. Hence, we have that $\tilde{\Psi}_{iL} = 1$ for every i.[12]

We provide some examples to illustrate how this notation can be used to represent common network structures.

Example 9 (Horizontal Economy)**.** A popular structure is a single CES consumption aggregator without input-output linkages. The economy has N producers, each of whom uses only labor to produce, and sells directly and exclusively to the household ($\Omega_{iL} = 1$, $\Omega_{ij} = 0$ for every $i \in N$). Household expenditure shares are given by Ω_{Ci} for $i \in N$ and $\Omega_{CC} = \Omega_{CL} = 0$ since the consumption sector does not purchase from itself or from labor. In this economy, the labor share is given by $\Lambda = \sum_j \frac{\Omega_{Cj}}{\mu_j}$, which is decreasing in markups. The economy has a horizontal structure and is depicted in Figure 1.

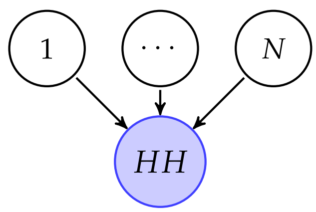

Figure 1: Horizontal economy. Arrows show flow of goods. Labor input is not shown.

Example 10 (Roundabout Economy)**.** The roundabout economy, popularized by Basu (1995), is the simplest way to introduce intermediate inputs. There is a single producer who buys intermediate inputs from itself. Denoting the intermediate input share in production by Ω_{11}, we have $\Omega_{1L} = 1 - \Omega_{11}$ and the labor share is given by $\Lambda = \frac{1}{\mu_1}$. This economy is depicted in Figure 2.

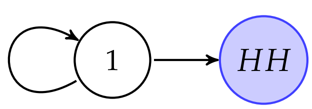

Figure 2: Roundabout economy. Arrows show flow of goods. Labor is not shown.

[12] Formally, the equality follows from the fact that cost shares must sum to one, $\sum_{j \in N} \tilde{\Omega}_{ij} + \tilde{\Omega}_{iL} = 1$. We can rewrite this equality as $\tilde{\Psi}_{iL} = \sum_j \tilde{\Psi}_{ij} \Omega_{jL} = 1$, where the first equality follows from the condition $\tilde{\Omega}_{Li} = 0$ for every $i \in \mathcal{I}$.

4 Microeconomic Propagation

Having defined the primitives of our sandbox economy, in this section we consider how technology and wedge shocks hitting one producer are transmitted to the prices and quantities of other producers. These results are of independent interest for understanding co-movement across the economy.[13] However, it turns out that, even if one is not interested in co-movement per se, aggregation in inefficient or nonlinear settings, studied in Sections 5 and 6, requires understanding how shocks propagate at a disaggregated level. Therefore, the results in this section will be useful when we turn our attention back to aggregation.

We begin by studying how prices respond to shocks, which we call forward propagation. We then consider how sales respond to shocks, which we call backward propagation. The log-difference between sales and price changes yields quantity changes. The results presented in this section are special cases of those in Baqaee and Farhi (2020c).

4.1 Prices: Forward Equations

Prices are determined through *forward propagation* equations. These equations describe how productivity and wedge shocks are pushed forward through supply chains from suppliers to their customers. To derive the forward equations, note that cost minimization by each producer implies that the price of $i \in \{1, \ldots, N\}$ is determined by

$$d \log p_i = \sum_{j=1}^{N} \mu_i \Omega_{ij} d \log p_j + \mu_i \Omega_{iL} d \log w + d \log \mu_i - d \log A_i. \tag{8}$$

In words, the price of i responds to changes in the price of each of its inputs in proportion to the share of that input in i's costs. Whenever there are non-unit markups, there is a gap between expenditures on each input as a share of revenues, Ω_{ij}, and expenditures as a share of costs, $\mu_i \Omega_{ij}$. One input is labor, whose share in total costs is $\mu_i \Omega_{iL}$. Finally, the price of i also responds to its own wedge $d \log \mu_i$ and technology $d \log A_i$. Equation (8) is

[13] Since our sandbox economy has a single factor of production and a representative agent, it is unsuitable for studying questions related to the distribution of income across different factors (e.g. labor and capital). However, in Section 8, we discuss how these results can be generalized to contexts with heterogenous consumers and multiple factors.

a system of linear equations in $d\log p$. Solving through for $d\log p$, we get[14]

$$d\log p = \mathbf{1} d\log w + \tilde{\Psi}\left(d\log\mu - d\log A\right), \tag{9}$$

Without loss of generality, we set nominal output to be the numeraire. Under this normalization, changes in the nominal wage are driven either by changes in the Domar weight of labor Λ or changes in the endowment L:

$$d\log w = d\log\Lambda - d\log L.$$

For now, we take the quantity of labor to be exogenously determined and fixed (so that $d\log L = 0$). Under these assumptions, we can rewrite (9) in the following way.

Proposition 1 (Forward Propagation)**.** *The change in prices in response to technology and wedge shocks is*

$$d\log p_i = \sum_{j=1}^{N} \tilde{\Psi}_{ij}\left(d\log\mu_j - d\log A_j\right) + d\log\Lambda. \tag{10}$$

In words, a wedge or technology shock propagates from upstream suppliers to downstream consumers through prices. The propagation is mediated by the Leontief inverse $\tilde{\Psi}$. In addition, all prices are equally affected by a change in the wage, which given our choice of numeraire, is just the change in labor's share of income. This formulation of the forward equation naturally extends to economies with multiple factors and consumers (for example labor and capital, or high- and low-skill workers).

Example 11 (Amplification of Productivity Shocks)**.** To illustrate how input-output linkages can amplify shocks, consider the roundabout economy depicted in Figure 2. Assume marginal cost pricing, $\mu = 1$. In the absence of wedges, labor's share of income is identically equal to one, and hence $d\log\Lambda = 0$. The forward equation (10) then pins down the price as a function of changes in productivity:

$$d\log p = -\frac{1}{1-\Omega_{11}} d\log A, \tag{11}$$

The effect of productivity on prices is increasing in the intermediate input share Ω_{11}. Intuitively, in this economy, a productivity shock lowers the price of the final good directly,

[14]To derive equation (9) we used the equality $\sum_j \tilde{\Psi}_{ij}\tilde{\Omega}_{jL} = 1$ for every $i \in \{1,\dots,N\}$, derived in Footnote 12.

but also indirectly by lowering the price of intermediates, intermediates' intermediates, and so on, ad infinitum. The cumulative effect of the productivity shock is given by $\sum_{t=0}^{\infty} \Omega_{11}^t = (1 - \Omega_{11})^{-1}$.

4.2 Sales: Backward Equations

The forward equations pin down changes in prices as a function of shocks and changes in sales (specifically, the sales share of labor). We now derive the backward equations, which pin down changes in sales (including the sales share of labor) as a function of changes in prices and shocks. Combining the backward and forward equations will yield a complete characterization of propagation patterns in terms of primitives.

The backward equations describe how changes in demand are transmitted backwards from buyers to suppliers. Like in simpler models, demand elasticities play a key role. When the relative price of some good i increases (due to changes in productivity and/or wedges), the expenditure share on i also increases if, and only if, i is complementary with the other goods in the economy, according to an appropriate notion of complementarity implied by the production network.

To derive the backward propagation equations, we differentiate the market clearing equation (7):

$$d\lambda' = d\lambda'\Omega + \lambda' d\Omega = \lambda' d\Omega \Psi. \tag{12}$$

We proceed term-by-term. For expositional clarity, suppose that each i has a CES production function with elasticity of substitution θ^i. In this case, changes in the expenditure shares of each i on j is

$$\begin{aligned} d\log \Omega_{ij} &= \left(1-\theta^i\right)\left(d\log p_j - \sum_k \tilde{\Omega}_{ik} d\log p_k\right) - d\log \mu_i \\ &= (1-\theta^i) Cov_{\tilde{\Omega}_{(i,:)}}\left(d\log p, I_{(:,j)}\right) - d\log \mu_i. \end{aligned} \tag{13}$$

The first line shows that i switches expenditures towards j if the price of j rises by more than prices on average, and if j is complements with other inputs. Furthermore, holding input prices constant, an increase in i's markups reduces expenditures on all inputs. The second line rewrites the change in a more compact way using an input-weighted covariance where $d\log p$ is a vector of price changes, $I_{(:,j)}$ is the jth column of the identity matrix, and the covariance uses the weights given by i's input shares $\tilde{\Omega}_{(i,:)}$. Substituting

(13) into (12), and using the distributive properties of covariances, yields the backward propagation equations in Proposition 2 below.

Proposition 2 (Backward Propagation)**.** *The change in the sales share of each good i is*

$$d\lambda_i = -\sum_{j=1}^{N} \lambda_j \Psi_{ji} d\log \mu_j + \sum_{k=1}^{N} \frac{\lambda_k}{\mu_k} \left(1-\theta^k\right) Cov_{\tilde{\Omega}_{(k,:)}} \left(d\log p, \Psi_{(:,i)}\right), \tag{14}$$

where $\tilde{\Omega}_{(k,:)}$ denotes the kth row of $\tilde{\Omega}$ and $\Psi_{(:,i)}$ denotes the ith column of Ψ. When specialized to labor, this is

$$d\Lambda = -\sum_{j=1}^{N} \lambda_j \Psi_{jL} d\log \mu_j + \sum_{k=1}^{N} \frac{\lambda_k}{\mu_k} \left(1-\theta^k\right) Cov_{\tilde{\Omega}_{(k,:)}} \left(d\log p, \Psi_{(:,L)}\right). \tag{15}$$

The intuition for both equations is similar, so focus on equation (15). The first set of summands captures a level effect, and tells us that the labor share falls if j's markup increases, and the importance of j depends on its size λ_j times the fraction of its revenues that are ultimately paid out to labor Ψ_{jL}. The second set of summands capture a composition effect, coming from expenditure-switching due to substitution. If producer k substitutes towards inputs with high labor payments, then this raises labor's share of income. This is the case when goods are substitutes ($\theta^k > 1$), and prices fall more for goods with high markups along their supply chain. That is, when goods are substitutes and the covariance between price changes $d\log p$ and total labor payments $\Psi_{(:,L)}$ is negative. Similarly, labor's share of income increases when good are complements and the covariance is positive (and vice versa).

Example 12 (Markups and the Labor Share in the Horizontal Economy)**.** In this example, we specialize equation (15) to the horizontal economy depicted in Figure 1. We restrict attention to changes in markups, so that $d\log p = d\log\mu + d\log\Lambda$. In this economy, $Cov_{\tilde{\Omega}_{(j,:)}}\left(d\log p, \frac{1}{\mu}\right) \neq 0$ only for $j = C$ (i.e. for the final consumer). Furthermore, $Cov_{\tilde{\Omega}_{(C,:)}}\left(d\log p, \frac{1}{\mu}\right) = Cov_{\tilde{\Omega}_{(C,:)}}\left(d\log \mu, \frac{1}{\mu}\right)$. Putting this all together, the change in the labor share is

$$d\Lambda = -\sum_{j=1}^{N} \lambda_j \mu_j^{-1} d\log\mu_j + (1-\theta^C) Cov_{\tilde{\Omega}_{(C,:)}}\left(d\log\mu, \frac{1}{\mu}\right). \tag{16}$$

The first term is a mechanical effect capturing the fact that, holding fixed initial spending, a change in markups changes the labor share. The second term describes the composition effect whereby spending patterns change in response to changes in relative prices. Specifically, if goods are complements, and relative prices rise for goods with initially high markups, as captured by the covariance $Cov_{\tilde{\Omega}_{(C,:)}}\left(d\log\mu, \frac{1}{\mu}\right)$, then the labor share falls. This is because spending shifts in favor of high-markup producers, and this further reduces the labor share.

Combining the backward equation (14) with the forward equation (10) pins down all prices, expenditures, and quantities.[15] In Sections 5 and 6 we see that these backward and forward equations are needed in order to understand aggregation in inefficient and nonlinear economies.

5 Aggregation with Inefficiencies

We now generalize the aggregation result in Theorem 1 beyond the efficient case for the sandbox economy.[16] To generalize Theorem 1, we define the allocation rule $\mathcal{X}$ of the economy to be the matrix that records the share of each good $j \in \mathcal{N}$ sent to each user $i \in \mathcal{N}$. We first present a non-parametric result, where changes in real GDP and the Solow residual are expressed as a function of the labor share. Corollary 1 then derives a parametric version of the result, using the forward and backward equations to express changes in the labor share in terms of primitives.

Proposition 3 (Real GDP and Solow Residual in Sandbox Economy)**.** *Given changes in technology $d\log A$, wedges $d\log\mu$, and factor quantities $d\log L$, the change in real GDP in the sandbox economy is*

$$d\log Y_{RGDP} = \underbrace{\sum_{i=1}^{N}\tilde{\lambda}_i d\log A_i + d\log L}_{\Delta\ technology}\underbrace{-\sum_{i=1}^{N}\tilde{\lambda}_i d\log\mu_i - d\log\Lambda,}_{\Delta\ allocation} \tag{17}$$

and the change in the Solow residual is

$$d\log Y_{SR} = \underbrace{(1-\Lambda)d\log L}_{\Delta\ external\ inputs} + \underbrace{\sum_{i=1}^{N}\tilde{\lambda}_i d\log A_i}_{\Delta\ technology}\underbrace{-\sum_{i\in N}\tilde{\lambda}_i d\log\mu_i - d\log\Lambda}_{\Delta\ allocation}. \tag{18}$$

[15] As shown by Baqaee and Farhi (2019a), repeatedly iterating on the forward and backward equations can also be used to compute exact (nonlinear) comparative statics.

[16] See Appendix B for a version of the theorem which applies to the more general setup in Section 2.

To derive (17), note that real GDP can be expressed as the change in nominal output deflated by an appropriately defined price index

$$d \log Y_{RGDP} = d \log \sum_{i \in \mathcal{N}} p_i y_i - \sum_{i \in \mathcal{N}} \frac{p_i y_i}{\sum_j p_i y_i} d \log p_i.$$

Plugging the forward equations (10) into the expression above yields (17).

Note that when prices are equal to marginal cost in the initial equilibrium, Proposition 3 collapses to Theorem 1. This is because when $\mu_i = 1$, we have that $\tilde{\lambda}_i = \lambda_i = p_i x_i / E$. Furthermore, if $\mu_i = 1$ for every i, then $d \log \Lambda = -\sum_i \lambda_i d \log \mu_i$ and $\Lambda = \tilde{\Lambda} = 1$.

Before explaining the intuition for Proposition 3, a word about its applicability. We treat changes in external inputs, productivity, and wedges as exogenous primitives. Theories of innovation, market frictions, consumption-leisure choice, and international trade endogeneize l, A, and μ in different ways. In these cases, l, A, and μ are themselves functions of some deeper primitive state variable. The determination of these functions could be complex and interesting in and of itself. In those cases, Proposition 3 still applies, but it must be supplemented with additional equations to conduct counterfactuals, since external inputs, technologies, and wedges may move simultaneously in response to a given shock. For example, equation (17) implies that reductions in the labor share are associated with gains in allocative efficiency. However, reducing the labor share may also cause employment to fall. In Section 7.1, we show how to endogenize labor supply, and in Section 7.2 we show one way of endogenizing wedges (sticky prices).

Now we discuss intuition. Proposition 3 spells out the three sources of aggregate changes discussed in Section 2: changes in external inputs, technology effects, and reallocation effects. We discuss the three forces highlighted by Proposition 3 in sequence. First, consider the external inputs effect. In contrast to the efficient case in Theorem 1, away from an efficient equilibrium changes in external inputs directly affect aggregate real activity. Following the definitions in Table 1, this effect does not appear for real GDP, since there are no external inputs. For the Solow residual, instead, labor is an external input, and, if the labor share Λ is less than one, then an increase in labor shows up as a positive change (as observed by Hall, 1990). Intuitively, the labor share is a measure of the aggregate inverse markup,[17] so that, when $\Lambda < 1$, the marginal revenue product of labor holding fixed the allocation rule exceeds its cost. Hence, as the quantity of labor increases, output increases proportionately to the marginal revenue product of labor, while

[17] See Section 3.2 for a discussion.

the Solow residual subtracts the additional labor input from the net output measure according to its cost. In the absence of wedges, this term is equal to zero since the labor share is identically equal to one.

Next, consider the direct effect of technology shocks. Similar to Hulten's theorem, the contribution of technology shocks in each sector to aggregate real activity is proportional to the relevant Domar weight. The only difference relative to Hulten's theorem is that we must use cost-based Domar weight, which adjust the traditional sales-based Domar weights for the presence of the wedges μ. In the absence of wedges, $\tilde{\lambda} = \lambda$, but with positive markups, the cost-based Domar weights will exceed the revenue-based Domar weights ($\tilde{\lambda} \geq \lambda$).

Finally, consider the reallocation effects in (17). There are beneficial reallocations if labor's share of income declines faster than $\tilde{\lambda}$ weighted wedges increase. Intuitively, beneficial reallocations are those that redirect resources towards the parts of the economy that were too small to begin with. Holding markups constant, the labor share falls if, and only if, firms with relatively heavily marked up supply chains get larger. That is, the labor share falls if firms whose sales are mostly dissipated as wedge income rather than labor income get larger. Since these are precisely the parts of the economy that are inefficiently too small, because wedges are high, the reduction in the labor share signals beneficial reallocation. If the wedges themselves are rising, then even holding the allocation of resources constant, the labor share will fall. The presence of the $-\sum_i \tilde{\lambda}_i d\log\mu_i$ term in (17) accounts for this mechanical (non-composition) effect of changes in wedges on the labor share.

In Corollary 1 we use the backward equation (15) to substitute for $d\log\Lambda$ in equation (17), providing an expression for the change in real GDP in terms of microeconomic primitives.

Corollary 1 (Real GDP in Terms of Primitives)**.** *Substituting for the labor share from the backward equation* (15) *yields*

$$d\log Y_{RGDP} = \underbrace{\sum_i \tilde{\lambda}_i d\log A_i + d\log L}_{\text{technology}} - \underbrace{\sum_j \frac{\lambda_j}{\mu_j}\left(\theta_j - 1\right) Cov_{\tilde{\Omega}_{(j,:)}}\left(\sum_i \tilde{\Psi}_{(:,i)} d\log A_i, \Psi_{(:,L)}\right)}_{\text{reallocation caused by technology}}$$

$$+\underbrace{\sum_j \frac{\lambda_j}{\mu_j}\theta_j Cov_{\tilde{\Omega}_{(j,:)}}\left(\sum_i \tilde{\Psi}_{(:,i)} d\log\mu_i, \Psi_{(:,L)}\right)}_{\text{reallocation caused by wedges}}. \tag{19}$$

The reallocation effects of productivity and wedge shocks are captured by the covariance between relative price changes, $d\log p - d\log\Lambda = \sum_i \tilde{\Psi}_{(:,i)}(d\log\mu_i - d\log A_i)$, and the cumulative inverse markups $\Psi_{(:,L)}$.[18] As usual, beneficial reallocations take place when resources shift towards producers with relatively high initial wedges, because from a social perspective these producers received too few resources to begin with.

Start by considering reallocation effects due to technology shocks and consider some producer j in the sum. A negative covariance means that suppliers of j with heavily marked-up supply chains, measured by $\Psi_{(:,L)}$, are also the ones whose relative prices rose due to the productivity shocks, measured by $\sum_i \tilde{\Psi}_{(:,i)} d\log A_i$. If j's inputs are substitutes, $\theta_j > 1$, then j will substitute away from these suppliers. In this situation, substitution by j redirects resources away from relatively high-markup parts of the economy, which lowers aggregate output since those parts of the economy were too small to begin with.

Next, consider the reallocation effects due to changes in wedges. Increasing i's wedge lowers i's demand for inputs. Therefore, resources are always (weakly) reallocated away from it. This reallocation is stronger the more substitutable are inputs. Therefore reallocation effects due to wedges are negative when increases in wedges positively covary with initial cumulative inverse markups, captured by $\Psi_{(:,L)}$, thereby increasing the overall dispersion across sectoral wedges.

To make this more concrete, consider the following example.

Example 13 (Reallocation in Horizontal Economy)**.** The horizontal economy clearly illustrates how shocks interact with initial wedges. Suppose that producers have heterogeneous initial markups μ_i, and are hit by a vector of productivity and wedge shocks $(d\log A, d\log\mu)$. Consumers have CES preferences, with elasticity of substitution θ^C. Equation (19) becomes

$$d\log Y_{RGDP} = \underbrace{\mathbb{E}_\lambda[d\log A]}_{\text{technology}} + \underbrace{\left(1-\theta^C\right) Cov_{\Omega_{(C,:)}}\left(d\log A, \frac{1}{\mu}\right)}_{\text{reallocation caused by technology}} + \underbrace{\theta^C Cov_{\Omega_{(C,:)}}\left(d\log\mu, \frac{1}{\mu}\right)}_{\text{reallocation caused by wedges}},$$

where $\lambda = \Omega_{(C,:)}$ are the household's consumption shares in the initial equilibrium, and

[18] Section 3.2 illustrates the relation between inverse markups and $\Psi_{(:,L)}$.

$\mathbb{E}_\lambda(\cdot)$ is the sales-weighted average. Consider reallocations due to productivity shocks. Allocative efficiency improves either if goods are substitutes and productivity increases for high-markup producers, or if goods are complements and productivity increases for low-markup producers. Either way, beneficial reallocations due to productivity shocks to take place if labor's share of income falls, as in Proposition 3.

Now consider reallocations due to changes in wedges. Allocative efficiency rises if markups fall for producers with initially high markups. This holds regardless of the value of the elasticity of substitution, but the effect is stronger if the elasticity is higher, since this implies stronger beneficial reallocations in response to the changes in markups. When the elasticity of substitution $\theta^C = 0$, then the change in labor's share of income, $d \log \Lambda$, is exactly equal to the mechanical change in the average wedges $-\mathbb{E}_\lambda[d \log \mu]$. In this case, there are no reallocations due to changes in wedges. When the elasticity of substitution is greater than zero, reallocations take place according the covariance of the change in the wedge and the initial wedge.

Given the example above, it is tempting to conclude that firms are too small from a social perspective if their markup/wedge is higher than average. In the misallocation literature, the wedge μ is called revenue productivity (TFPR). Most models of misallocation have a horizontal structure, as in Figure 1, and so firms with higher TFPR (wedges) are too small and ones with lower TFPR are too large relative to first-best. The next example shows that when firms are interdependent, this logic breaks down.

Example 14 (Reallocations with Interdependent Firms)**.** Consider the economy depicted in Figure 3. Firm 1 produces directly from labor but firm 2 sources its input from firm 3, and firm 3 uses labor. Suppose that $\mu_1 > \mu_3 = 1$. That is, producer 1 has a higher markup than producer 3. Firm 1 and 3 are the only users of labor, and since 1 has the higher markup, one might naively expect that reallocating workers from 3 to 1 should raise aggregate efficiency. After all, in the jargon of the literature on misallocation, producer 1 has a higher revenue productivity (TFPR) than firm 3, and these two firms are the only ones competing over workers in this economy. To achieve such a reallocation, consider a subsidy τ_1 for 1's use of labor. This is isomorphic to a reduction in 1's markup. Corollary 1 implies that

$$\frac{d \log Y}{d \log \tau_1} = -\theta_0 \lambda_1 (1 - \lambda_1) \bar{\mu} \left[\frac{\mu_2 - \mu_1}{\mu_1 \mu_2} \right] d \log \tau_1.$$

That is, the subsidy will reduce output if $\mu_2 > \mu_1$, despite the fact that 2 is not directly using any labor. This example underscores that when firms are interdependent, one cannot compare wedges to wedges, in this case μ_1 to μ_3, in isolation to determine the consequences of reallocations. This harkens back to the classic arguments of McKenzie (1951) who cautioned against this type of logic in designing policies.

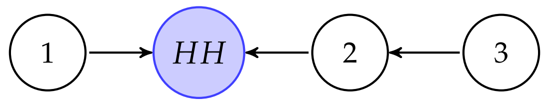

Figure 3: Arrows direct the flow of goods. The primary factor input (labor) is not shown.

6 Nonlinear Aggregation

As mentioned earlier, the forward and backward equations are also a building block to study the nonlinear effect of productivity and wedges on aggregate output. For instance, shocks to small producers can have large aggregate effects due to nonlinearities and complementarities in production.

In this section, we use the forward and backward equations to characterize, to a second-order, the aggregate consequences of technology shocks and distortions. Consider the second-order approximation

$$\Delta \log Y \approx \frac{\partial \log Y}{\partial \log A} \cdot \Delta \log A + \frac{\partial \log Y}{\partial \log \mu} \cdot \Delta \log \mu + \frac{1}{2} \nabla^2 \log Y,$$

around $\mu = \mathbf{1}$. The first-order terms are given by Theorem 1, so we focus on the second-order terms, which can be spelled out as

$$\nabla^2 \log Y = d \log A^T \frac{\partial^2 \log Y}{\partial \log A^2} d \log A + 2 d \log A^T \frac{\partial^2 \log Y}{\partial \log A \partial \mu} d \log \mu + d \log \mu^T \frac{\partial^2 \log Y}{\partial \log \mu} d \log \mu.$$

The terms in this equation capture nonlinearities with respect to technology shocks and wedges. The first set of summands are nonlinearities associated with technology shocks holding fixed wedges, the second set of summands are nonlinearities associated with interactions between technology and wedge shocks (cross-partial derivatives between wedge and technology shocks), and the last set of summands are nonlinearities associated with wedges holding fixed technology.

The following two propositions, extending Baqaee and Farhi (2020c), describe nonlinearities associated with technology and wedge shocks. Proposition 4 expresses aggregate nonlinearities in terms of (endogenous) changes in quantities and expenditure shares induced by the shocks. Proposition 5 then re-expresses this result in terms of microeconomic primitives using the additional structure provided by our sandbox economy.

Proposition 4 (Nonlinear Aggregation Near Efficiency)**.** *Consider the economy described in Section 2. Suppose the supply of external inputs is fixed. Around the efficient allocation, up to a second-order approximation in* $(d\log A, d\log \mu)$*, the change in aggregate output is*

$$\nabla^2 \log Y = \underbrace{\sum_{i,j} \lambda_i \frac{\partial \log \lambda_i}{\partial \log A_j} d\log A_j d\log A_i}_{\text{technology shocks}} + \underbrace{\sum_{i,j} \lambda_i \frac{\partial \log x_i}{\partial \log \mu_j} d\log \mu_j d\log \mu_i}_{\text{wedge shocks}}, \tag{20}$$

assuming that $\Delta \log Y$ *is path-independent and* A_i *are Hicks-neutral.*[19]

Equation (20) shows that the nonlinear effect of subjecting the economy to technology and wedge shocks can be broken down into two separate terms. Specifically, at the efficient allocation, there are no interactions (cross-partial derivatives) between wedge and technology shocks. Furthermore, nonlinearities due to technology shocks can be written in terms of changes in expenditure shares caused by technology shocks. On the other hand, nonlinearities due to wedges can be written in terms of changes in quantities caused by wedges.

The changes in quantities and expenditure shares in equation (20) are endogenous. Using the forward and backward propagation equations (10) and (14), Proposition 5 rewrites them in terms of the microeconomic primitives instead.

Proposition 5 (Nonlinear Aggregation in Terms of Primitives)**.** *For the sandbox economy, around the efficient allocation, the second-order derivative of real GDP with respect to productivity and wedges is given by*

$$\nabla^2 \log Y_{RGDP} = \underbrace{-\sum_{j\in\mathcal{N}} \lambda_j \left(1-\theta^j\right) Var_{\Omega_{(j,:)}}\left(\sum_i \Psi_{(:,i)} d\log A_i\right)}_{\text{technology shocks}} \underbrace{-\sum_j \lambda_j \theta^j Var_{\Omega_{(j,:)}}\left(\sum_i \Psi_{(:,i)} d\log \mu_i\right)}_{\text{wedge shocks}}.$$

[19] As mentioned before, the assumption of Hicks neutrality is easy to generalize simply by replacing $d\log A_i$ with $(\partial \log f_i/\partial \log A_i) d\log A_i$ instead. The assumption that $\Delta \log Y$ is path-independent, however, is not without loss of generality. Throughout the rest of this section, we assume that the integral in (3) is path-independent. This is ensures $\Delta \log Y$ can be written as a function of the initial and final state only and does not depend on the path of integration. In the sandbox economy of Section 3, real GDP and the Solow residual satisfy this requirement. See Hulten (1973) and Baqaee and Burstein (2021) for more information on path-dependence.

The first part gives the nonlinear effects of technology shocks—these terms can be negative or positive depending on whether aggregate output is concave or convex in the technology shocks. The second part is the nonlinear effects of wedge shocks—these terms are always negative since the introduction of wedges into an efficient economy necessarily reduces output.

Since there are no interactions between productivity and wedge shocks in Proposition 5, we discuss the effect of productivity and wedges in isolation below, starting with technology shocks.

Nonlinear Effect of Technology Shocks. Denoting the (unapproximated) change in real GDP by $\Delta \log Y_{RGDP}$, and focusing our discussion on technology shocks, assuming away wedges, Proposition 5 implies the following second-order approximation:

$$\Delta \log Y_{RGDP} \approx \sum_i \lambda_i \Delta \log A_i - \frac{1}{2} \sum_j \lambda_j \left(1 - \theta^j\right) Var_{\Omega_{(j,:)}} \left(\sum_i \Psi_{(:,i)} \Delta \log A_i \right), \qquad (21)$$

where $\Delta \log A_i$ is the productivity shock to i. The sign of the second order terms in equation (21) depends on whether the microeconomic elasticities of substitution θ^j are greater than or less than one. Intuitively, productivity changes generate price dispersion. This is beneficial when goods are substitutes, because producers and consumers can shift their expenditures towards cheaper goods. By contrast, when goods are complements, output is constrained by the most expensive (least productive) input.

Recall from the forward equation (10) that the term $\sum_i \Psi_{(:,i)} \Delta \log A_i$ corresponds to the change in the price of each good relative to labor in response to the vector of productivity shocks $\Delta \log A$. Producer j will substitute its expenditures across inputs in accordance to the price dispersion it faces in the price of its inputs and how different its elasticity of substitution is from one. The relevant price dispersion for each producer j is given by the variance of relative prices weighted by its input shares $Var_{\Omega_{(j,:)}} \left(\sum_i \Psi_{(:,i)} \Delta \log A_i \right)$. This expenditure-switching by j changes the Domar weights in the economy. The larger is j's own Domar weight, λ_j, the more expenditure-switching by j will affect the Domar weight of other producers. By Proposition 4, through a nonlinear effect, changes in productivity have a larger impact on aggregate output if the Domar weights increase for the sectors whose productivity also increased. To illustrate the intuition, consider the following example.

Example 15 (Nonlinearity due to Critical Inputs)**.** Consider an economy with a star supplier depicted in Figure 4. There are N retailers who sell directly and exclusively to the household. Each retailer i relies on a mix of labor and an intermediate input labeled e (say energy). Energy does not use materials, so $\Omega_{Ei} = 0$, for every i. Retailers do not use materials other than energy, so $\Omega_{ij} = 0$ for $j \neq e$. Finally, energy uses only the primary factor $\Omega_{eL} = 1$, whereas the labor intensity of retailers is inversely related to energy intensity $\Omega_{iL} = 1 - \Omega_{ie}$.

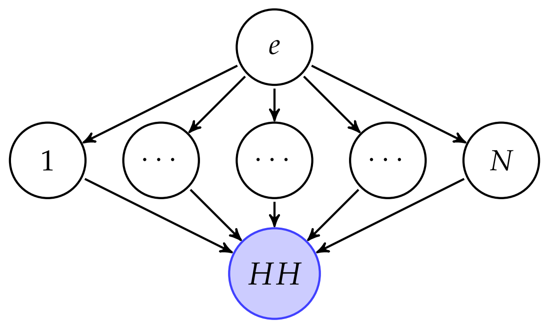

Figure 4: Star input economy. The arrows direct the flow of goods. The primary factor input (labor) is not shown.

Suppose the star input (energy) and labor are complements in production, but consumption goods are substitutes: $\theta^j < 1 < \theta^C$. Furthermore, suppose that there are no productivity shocks to the retailers ($\Delta \log A_i = 0$), while energy productivity changes by $\Delta \log A_e$. Applying (21) gives

$$\Delta \log Y_{RGDP} \approx \lambda_e \Delta \log A_e - \frac{1}{2} \left[\sum_j \left(1 - \theta^j\right) \lambda_j \Omega_{je} (1 - \Omega_{je}) + \left(1 - \theta^C\right) Var_{\Omega_{(C,:)}} \left(\Omega_{(:,e)}\right) \right] \Delta \log A_e^2.$$

Focus on the nonlinear terms. If inputs are complements ($\theta^i < 1$) the terms involving $1 - \theta^i$ are negative, because retailers' output is constrained by the least productive of the two inputs (energy and labor). If energy becomes more productive, then retailers are constrained by the fact that labor productivity has not increased, and vice versa. On the other hand, the term involving $1 - \theta^C$ is positive. If producers have heterogeneous energy intensity $\Omega_{e,i}$, and final goods are substitutes ($\theta^C > 1$), consumers increase their expenditure on goods with lower (higher) energy intensity after a negative (positive) energy shock. This diminishes the effect of negative shocks and amplifies the effect of positive ones.

Nonlinear Effect of Wedges. We now shift focus to the role of distortions and assume away technology shocks. In this case, Proposition 5 simplifies to

$$\Delta \log Y_{RGDP} \approx \frac{1}{2}\sum_i \lambda_i \Delta \log x_i \Delta \log \mu_i = -\frac{1}{2}\sum_j \lambda_j \theta_j Var_{\Omega_{(j,:)}}\left(\sum_i \Psi_{(:,i)} \Delta \log \mu_i\right), \qquad (22)$$

where $\Delta \log x_i$ and $\Delta \log \mu_i$ are the changes in wedges and total quantity i. By Theorem 1, the first-order terms are zero so there are only nonlinear terms. The first equality shows that the reduction in aggregate productivity caused by wedges is approximately equal to the Domar-weighted sum of deadweight-loss (Harberger 1964) triangles throughout the economy. These losses from misallocation are larger if Domar weights are larger and if quantities are more elastic to wedges. The second equality expresses these Harberger triangles in terms of the input-output matrix and elasticities of substitution.

Unlike with productivity shocks, where the second-order effects originating from each producer j are positive if, and only if, its inputs are substitutes, for wedges, the second-order terms are always (weakly) negative, and more so when goods are better substitutes. Intuitively, expenditure-switching by consumers and producers inefficiently reallocates quantities away from sectors with higher relative markup. This effect is stronger when elasticities of substitution, θ^j, are larger. Again, we turn to some examples to illustrate the intuition.

Example 16 (Misallocation in the Horizontal Economy)**.** Consider the horizontal economy in Figure 1. Applying (22) yields

$$\Delta \log Y_{RGDP} \approx -\frac{1}{2}\theta^C Var_\lambda \left(\Delta \log \mu\right).$$

This equation shows that markup dispersion across sectors lowers aggregate output, and more so for higher elasticities of substitution.[20] Consumers inefficiently substitute toward low-markup goods, so that labor is not allocated to the goods with highest marginal revenue product. The reallocations caused by the wedges, and hence the extent of misallo-

[20]This is reminiscent of the formula derived by Hsieh and Klenow (2009) under the assumption that markups and productivities are jointly lognormal. The difference is that the variance in this formula is weighted by sales, whereas the variance term in Hsieh and Klenow (2009) is unweighted.

cation, is larger when goods are better substitutes.

Example 17 (Amplification through Input-Output Linkages.)**.** Consider the roundabout economy depicted in Figure 2. An application of (22) yields

$$\Delta \log Y_{RGDP} \approx -\frac{1}{2}\theta \frac{\Omega_{11}}{(1-\Omega_{11})^2} \left(\Delta \log \mu_1\right)^2.$$

Losses are monotone in the elasticity of substitution θ and in the intermediate input share Ω_{11}. Following (22), the right-hand side in the equation above corresponds to the Domar weight λ_1 times the deadweight loss triangle $\frac{1}{2}\Delta \log x_1 \Delta \log \mu_1$, where x_1 is the total output and μ_1 is the wedge on producer 1. The Domar weight is $1/(1-\Omega_{11})$. Hence, a given deadweight loss triangle is more costly if the intermediate input share is higher because it raises the Domar weight. The change in x_1 caused by the wedge is $\theta\Omega_{11}/(1-\Omega_{11})$. Hence, the triangle is larger if labor and materials are more substitutable or if the intermediate input share is larger. This is because a higher intermediate input share corresponds to more rounds of intermediate production, and distortions between labor and materials are compounded at every round. That is, intermediate inputs amplify the misallocation losses by enlarging both the Harberger triangles themselves and the weights used to aggregate them.

7 Extensions

We end our analysis by discussing some extensions to endogenize employment and wedges.

7.1 Endogenous Employment

Connecting the input-output framework with the business cycles literature requires modeling the comovement between employment and productivity. To do so we introduce elastic labor supply, by adding disutility from labor into the representative consumer's preferences, along the lines in Bigio and La'O (2020).[21] The utility function becomes

$$U(C, L) = \frac{C^{1-\gamma}}{1-\gamma} - \frac{L^{1+1/\varphi}}{1+1/\varphi},$$

[21] For input-output models with labor-leisure choice, see for example, the real business cycle model of Long and Plosser (1983), the misallocation model of La'O and Bigio (2020), and the New Keynesian models of Nakamura and Steinsson (2010), Pasten et al. (2017), Rubbo (2020).

where γ controls the income effect on labor and φ is the Frisch elasticity of labor supply. We can combine the labor-leisure condition with the forward equation (10) to write

$$\gamma d\log C + \frac{1}{\varphi} d\log L = \sum_{i\in N} \tilde{\lambda}_i (d\log A_i - d\log \mu_i). \tag{23}$$

Furthermore, since $d\log Y_{RGDP} = d\log C$, this labor market equation can easily be combined with the forward, backward, and aggregation equations (Propositions 1, 2, and 3) to solve for all endogenous variables.

Equation (23) shows that markups can now affect output and welfare through a new channel. Not only can markups distort how resources are allocated across producers, they can now also distort the quantity of labor supplied. This tempers the beneficial effects of reallocating resources towards high-markup producers. While reallocating resources to high-markup firms can improve the cross-sectional allocation of resources holding employment fixed, it can also reduce equilibrium employment by raising average markups.

7.2 Endogenous Wedges via Sticky Prices

Our final extension shows how the framework developed in this paper can easily incorporate endogenous wedges, such as those caused by nominal rigidities. This section builds on the analysis by Rubbo (2020) and La'O and Tahbaz-Salehi (2019).[22] They show that properly accounting for the input-output structure has important implications for the Phillips curve, the degree of monetary non-neutrality, and the conduct of optimal policy. Specifically, the Phillips curve is flatter when producers use intermediate inputs, and the divine coincidence fails so that monetary policy cannot implement the first-best equilibrium.

To nest nominal rigidities in our framework, we must endogenize the wedges that we have so far treated as primitives. To do this, consider an economy where firms must set prices before they can observe nominal wages and productivity shocks, and only a fraction δ_i of the firms in each sector i can adjust their price in response to shocks (as in Blanchard and Kiyotaki, 1987). A "sticky" price is simply a wedge that moves in the opposite direction to the marginal cost. Thus all the results derived so far can be applied

[22] Basu (1995) pioneered the importance of accounting for intermediate inputs in monetary models. Carvalho (2006) pointed out the importance of heterogeneity in price-stickiness for monetary non-neutrality. Our analysis here also relates to Pasten et al. (2017), Pasten et al. (2018), and Castro (2019) who study the relationship between production networks and sticky prices.

to an economy with nominal rigidities, with the caveat that we need to solve for these endogenous pricing wedges.

With some abuse of notation, let δ denote the diagonal matrix whose ith diagonal element is δ_i. Price stickiness means that only a faction of changes in marginal costs are passed through to prices. For simplicity, assume that shocks are unexpected. So the change in the price of each sector is given by $d\log p = \delta d\log mc$, where mc is nominal marginal cost.[23] The wedges, gaps between prices and marginal costs, are therefore given by

$$d\log \mu = -\left(I-\delta\right) d\log mc.$$

Combining these two relationships yields a relationship between changes in wedges and changes in prices

$$d\log \mu = -\left(I-\delta\right) \delta^{-1} d\log p. \tag{24}$$

Suppose that the initial allocation is efficient and $\log \mu = 0$ in steady state. Denote by e_i the ith basis vector, that is a vector of all zeros and a value of 1 in the $i$the element. Substituting (24) into (10) yields a sticky-price version of the forward equation:

$$d\log p = \delta\left(I-\Omega\delta\right)^{-1} e_L \left(d\log w - d\log A\right). \tag{25}$$

This equation shows that the transmission of changes in the wage and changes in productivity is dampened by sticky prices. From Theorem 1,

$$d\log Y_{RGDP} = \sum_{i\in N} \lambda_i d\log A_i + d\log L. \tag{26}$$

Combining (25), (26), and the labor market condition (23) yields sector-by-sector Phillips curves, written in vector form, as

$$d\log p = \mathcal{B}\left(\gamma + \frac{1}{\varphi}\right)\left(d\log L - d\log L^*\right) + \mathcal{V} d\log A, \tag{27}$$

where $d\log L^*$ is the change in employment in a version of the model without sticky

[23] As explained in Section 3, we include final consumption and labor among the set of producers. Therefore, our setup allows for sticky wages and/or sticky prices of consumption retailers (if $\delta_L < 1$ or $\delta_C < 1$).

prices (often called the *natural* level of employment) and $\mathcal{B}$ and $\mathcal{V}$ are constants

$$\mathcal{B} \equiv \frac{\delta\left(I-\Omega\delta\right)^{-1}e_L}{1-e_C'\delta\left(I-\Omega\delta\right)^{-1}e_L}, \quad \mathcal{V} \equiv \delta\left(I-\Omega\delta\right)^{-1}\left[\frac{e_L\lambda'\left(I-\delta\right)\left(I-\Omega\delta\right)^{-1}}{1-e_C'\delta\left(I-\Omega\delta\right)^{-1}e_L}-I\right].$$

The sectoral Philips curve relates changes in prices, for each sector, to deviations of aggregate employment from its efficient level and productivity shocks. Since all components of the vector $\delta\left(I-\Omega\delta\right)^{-1}e_L$ decrease when producers use less labor and more intermediate inputs, more intermediate input usage flattens the sectoral Phillips curves.

Equation (27) shows that in the presence of productivity shocks, sectoral inflation rates are not stabilized even if employment is at the efficient level ($d\log L = d\log L^*$). This is true also of almost all averages of sectoral inflation rates (i.e. linear combinations of sectoral inflation) like the consumer price index. Therefore monetary policy faces a tradeoff between achieving the efficient level of employment and stabilizing consumer inflation — that is, the divine coincidence of Blanchard and Gali (2007) fails.

The interaction between productivity shocks and nominal rigidities, which prevent the necessary adjustments in relative prices across firms and sectors, generate an additional welfare loss compared to the one-sector New Keynesian model. The additional welfare losses is described by equation (22) and fundamentally changes the welfare implications. Monetary policy has only one instrument (usually, the nominal interest rate), therefore it is not able to correct relative prices, resulting in a large welfare loss compared to an economy with flexible prices even under the optimal policy.

8 Related literature and the Broader Picture

We conclude by providing a brief overview of some of the issues to which the framework above has been or can be applied. Our discussion highlights how the methods presented in this review provide a unified framework to address a wide variety of seemingly unrelated applied questions and ongoing topics of research.

Microeconomic Origins of Aggregate Flucutations. Several studies discuss the extent to which idiosyncratic shocks contribute to the variance of aggregate output. Gabaix (2011) argues that, when the distribution of firm sizes is fat-tailed, idiosyncratic shocks to the productivity of large firms are not diversified at the aggregate level. A fat-tailed firm size distribution may arise for many different reasons — asymmetric input-output linkages, as in Acemoglu et al. (2011) or Carvalho and Tahbaz-Salehi (2019), the dynamics of firm growth, as in Carvalho and Grassi (2019), or fixed costs, as in Magerman et al. (2016). Most of the papers connecting microeconomic shocks to aggregate business cycles via input-output linkages rely on analytical tools similar to those surveyed in Section 3, especially Hulten's theorem, which we generalized in Theorem 1.[24]

Transmission of Shocks. Empirical tests of the economic mechanisms described in the paper are provided by Acemoglu et al. (2016), who test simplified versions of the forward and backward propagation of demand and supply shocks in equations (10) and (14) using industry-level data. In a similar vein, but using more disaggregated data, Carvalho et al. (2016), Boehm et al. (2015), and Barrot and Sauvagnat (2016) use natural disasters to study the propagation of shocks between suppliers and their customers. These studies emphasize the importance of complementarities in production. More structural, model-based, analyses of comovement and propagation of shocks in input-output networks include Foerster et al. (2011) and Atalay (2017) at business cycle frequencies and Foerster et al. (2019) for long-run growth. Relatedly, Vollrath (2021) characterizes the elasticity of substitution between labor and capital at the aggregate level using the forward and backward equations.

Cross-sectional Misallocation. A large literature seeks to measure misallocation and its effects on economic development (e.g. see surveys Restuccia and Rogerson 2013, 2017). This literature suggests that misallocation is a possible driver of cross-country (Restuccia and Rogerson 2008, Hsieh and Klenow 2009, Jones 2011) or cross-region (Boehm and Oberfield 2020) income differences. To quantify misallocation in the data, and its impact on aggregate output, these studies adopt measures that are closely related to equation (22) and the analysis in Section 6. Relatedly, Baqaee and Farhi (2020c), Edmond et al. (2018), and Osotimehin and Popov (2020) study the welfare losses from markups in different contexts, accounting for the role of intermediates.

Time-series Misallocation. Misallocation can also affect growth rates of a given country over time. Some papers focus on misallocation over longer-run horizons. For example,

[24]Relatedly, some papers connect input-output linkages with higher-order moments of aggregate output like skewness and kurtosis, for example Acemoglu et al. (2017), Baqaee and Farhi (2019b), and Dew-Becker et al. (2021). These papers also rely on the type of tools we surveyed. See also Dew-Becker (2021) who studies tail risk in production networks using asymptotic, rather than Taylor, expansions.

using a version of Proposition 3, Baqaee and Farhi (2020c) show that reallocations across firms with differing markups can explain as much as half of observed aggregate productivity growth in the US. Liu (2019) considers the implications of wedge-distortions for the conduct of industrial policy in a Chinese context. In an open economy setting, Edmond et al. (2015) show how trade liberalization affects welfare and productivity, over time, by changing markups. Bau and Matray (2020) combine Proposition 4 with a quasi-experimental research design to quantify how FDI liberalization raised industry-level Solow residuals in Indian manufacturing. Gopinath et al. (2017) study misallocation in capital markets, and follow a similar logic to equation (19) to document an inefficient allocation of the capital inflow in Southern Europe during the Euro convergence period.

Misallocation can also vary at business cycle frequencies. For example, Baqaee et al. (2021) and Meier and Reinelt (2020) use versions of Proposition 3 to show that aggregate demand shocks can change aggregate productivity by inducing reallocations across producers in models with endogenous markups and sticky prices. Another example of a business cycle application is Bigio and La'O (2020), who study the effects of financial frictions, represented as exogenous firm-specific wedges on capital, during the Great Recession. Financial frictions, like borrowing constraints, can be represented as endogenous wedges. Characterizing these wedges and their impact on long- and short-run changes in output is a very promising avenue for future work.

Endogenous Market Power. For the most part, we treat markups as exogenous wedges (with the exception of Section 7.2). In practice, firms optimally adjust their markups in response to changes in costs or demand. Endogenous changes in markups in production networks have been studied by Grassi (2017) and Kikkawa et al. (2019). Relatedly, Burstein et al. (2020) show the importance of explicitly accounting for the level of aggregation when studying micro- and macro-level patterns of markup cyclicality.

Investment and Dynamics. Technically, a dynamic interpretation of the model in Section 3, with goods indexed by state and time nests many dynamic business cycle models (e.g. (Long and Plosser, 1983)). In practice, dynamic models have more structure, like block-recursion, that can be exploited to derive sharper theoretical predictions. For example, Liu and Tsyvinski (2020) exploit time-separability assumptions to show that input adjustment costs generate persistence — and more so when production chains have more stages. Using a more complex dynamic model, vom Lehn and Winberry (2021) document that technology shocks to investment-producing sectors have become increasingly important drivers of business cycle fluctuations. Characterizing and spelling out dynamic versions of the forward and backward propagation equations is another interesting and underexplored area for research.

Entry-Exit and Link-Formation. Our analysis assumes all quantities can adjust smoothly: in the model sketched in Section 2, input-output flows are captured by smoothly-changing input shares. Entry-exit decisions, as well as link-formation at the most granular levels, are often lumpy because they feature fixed costs. Working with models that are nonetheless smooth, despite having an extensive margin, Baqaee (2018) and Baqaee and Farhi (2020a) show how the framework discussed in this paper can accommodate fixed costs of entry and exit, as well as increasing returns to scale. Baqaee et al. (2022) empirically estimate the value of extensive-margin relationships in production networks and develop growth-accounting formulas to allow for churn in supply chains. A vibrant branch of this literature provides explicit models of buyer-supplier link formation, oftentimes allowing aggregate output to be non-smooth with respect to shocks (Chaney, 2014; Oberfield, 2018; Lim, 2017; Huneeus, 2018; Taschereau-Dumouchel, 2020; Kopytov et al., 2022; Elliott et al., 2022). Elliott and Golub (2022) provide a recent review of how non-smooth cascades and fragilities can affect aggregate functionality in economic networks.

Heterogeneous Agents Models. A growing literature on heterogeneous agents in macroeconomics focuses on heterogeneity in asset portfolios and borrowing constraints. These in turn create heterogeneity in marginal propensities to consume across individuals. A dynamic version of our framework can nest heterogenous agent models with incomplete financial markets. Borrowing constraints and market incompleteness lead to wedges between realized and efficient intertemporal prices. Although our framework is silent about the endogenous determination of these wedges, it does allow us to study their consequences for real activity and welfare. As mentioned above, characterizing this endogenous mapping and exploring this application is an exciting area for future work.

Relatedly, most macroeconomic models with heterogeneous agents focus on the consumption-saving problem, but assume an exogenous distribution of labor income and a common consumption basket. Some recent papers have investigated heterogeneity in labor income and consumption prices arising from agents' differential exposure to different producers (see, e.g., Baqaee and Farhi, 2018; Guerrieri et al., 2020; Baqaee and Farhi, 2020b;

Flynn et al., 2022; Rubbo, 2022). Specifically, different agents are employed in different sectors, and consume different goods. If sectors face heterogeneous shocks, or if they are differentially affected by common shocks, heterogeneity across producers can cause heterogeneity in the agents' real income. Moreover, to the extent that agents with different asset portfolios also consume different goods, heterogeneity in financial income can affect heterogeneity in labor income. Characterizing this interaction is also a promising avenue for future research.

International Economics. The answers to classic questions in international economics, like the gains from trade, are highly sensitive to how the production structure is modelled. For example, in their handbook chapter, Costinot and Rodriguez-Clare (2014) show that the gains from trade can increase dramatically once one accounts for the presence of input-ouput connections. Although we abstract from international trade for most of this paper, Baqaee and Farhi (2019a) show how the framework we presented can be extended into an open-economy setting. Atkin and Donaldson (2021) use this extended framework to synthesize the way trade policy affects welfare in developing economies, accounting for the many different and interacting types of misallocation these economies suffer from. Kleinman et al. (2021) and Arkolakis et al. (2021) extend the type of models studied here to account for forward-looking migration, investment, and search-and-matching frictions. Meanwhile, researchers have used firm-to-firm VAT data to understand the transmission of international trade shocks through production networks at the individual firm-level (for example, Dhyne et al., 2022 and Bernard and Moxnes, 2018). This is also an exciting and evolving research area.[25]

References

Acemoglu, D., U. Akcigit, and W. Kerr (2016, January). Networks and the Macroeconomy: An Empirical Exploration. *NBER Macroeconomics Annual 30*, 273–335. Publisher: The University of Chicago Press.

[25] A parallel literature in international macroeconomics documents that fluctuations in aggregate output and inflation are correlated across countries, and — using propagation equations similar to the ones presented in Section 4 — investigates the importance of common shocks versus the transmission of idiosyncratic shocks via trade linkages in creating this correlation (e.g. di Giovanni et al., 2018 or Huo et al., 2019). More broadly, heterogeneous agent versions of the framework surveyed in this paper, along the lines described above, could be used to model international comovement, where agents correspond to consumers and workers in different countries.

Acemoglu, D., V. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2011). The Network Origins of Aggregate Fluctuations. *Econometrica 80*.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2017, January). Microeconomic Origins of Macroeconomic Tail Risks. *American Economic Review 107*(1), 54–108.

Arkolakis, C., F. Huneeus, and Y. Miyauchi (2021). Spatial production networks. *Unpublished, Yale University*.

Atalay, E. (2017, October). How Important Are Sectoral Shocks? *American Economic Journal: Macroeconomics 9*(4), 254–280.

Atkin, D. and D. Donaldson (2021). The role of trade in economic development. Technical report, National Bureau of Economic Research.

Bachmann, R., D. Baqaee, C. Bayer, M. Kuhn, A. Löschel, B. Moll, A. Peichl, K. Pittel, M. Schularick, et al. (2022). What if? the economic effects for germany of a stop of energy imports from russia. Technical report.

Baqaee, D., A. Burstein, C. Duprez, and E. Farhi (2022). Supplier churn and growth: a micro-to-macro analysis. Technical report.

Baqaee, D. and A. T. Burstein (2021). Welfare and output with income effects and taste shocks. *NBER Working Paper* (w28754).

Baqaee, D. and E. Farhi (2019a, July). Networks, Barriers, and Trade. Working Paper 26108, National Bureau of Economic Research. Series: Working Paper Series.

Baqaee, D. and E. Farhi (2020a, May). Entry vs. Rents: Aggregation with Economies of Scale. Technical Report w27140, National Bureau of Economic Research, Cambridge, MA.

Baqaee, D. and E. Farhi (2020b, May). Supply and Demand in Disaggregated Keynesian Economies with an Application to the Covid-19 Crisis. Working Paper 27152, National Bureau of Economic Research. Series: Working Paper Series.

Baqaee, D., E. Farhi, and K. Sangani (2021, January). The Supply-Side Effects of Monetary Policy. Technical Report w28345, National Bureau of Economic Research.

Baqaee, D. R. (2018). Cascading Failures in Production Networks. *Econometrica 86*(5), 1819–1838.

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

Baqaee, D. R. and E. Farhi (2019b). The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem. *Econometrica 87*(4), 1155–1203.

Baqaee, D. R. and E. Farhi (2020c). Productivity and Misallocation in General Equilibrium. *The Quarterly Journal of Economics 135*(1), 105–163. Publisher: Oxford University Press.

Barrot, J.-N. and J. Sauvagnat (2016, August). Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks. *The Quarterly Journal of Economics 131*(3), 1543–1592. Publisher: Oxford Academic.

Basu, S. (1995). Intermediate Goods and Business Cycles: Implications for Productivity and Welfare. *The American Economic Review 85*(3), 512–531. Publisher: American Economic Association.

Basu, S. and J. G. Fernald (1997). Returns to Scale in U.S. Production: Estimates and Implications. *Journal of Political Economy 105*(2), 249–283. Publisher: The University of Chicago Press.

Basu, S., L. Pascali, F. Schiantarelli, and L. Serven (2022, January). Productivity and the Welfare of Nations. *Journal of the European Economic Association*, jvac002.

Bau, N. and A. Matray (2020, October). Misallocation and Capital Market Integration: Evidence From India. Technical Report w27955, National Bureau of Economic Research, Cambridge, MA.

Bernard, A. B. and A. Moxnes (2018). Networks and Trade. *Annual Review of Economics 10*(1), 65–85. _eprint: https://doi.org/10.1146/annurev-economics-080217-053506.

Bigio, S. and J. La'O (2020, November). Distortions in Production Networks*. *The Quarterly Journal of Economics 135*(4), 2187–2253.

Blanchard, O. and J. Gali (2007). Real Wage Rigidities and the New Keynesian Model. *Journal of Money, Credit and Banking 39*(s1), 35–65. Publisher: Blackwell Publishing.

Blanchard, O. J. and N. Kiyotaki (1987). Monopolistic Competition and the Effects of Aggregate Demand. *The American Economic Review 77*(4), 647–666. Publisher: American Economic Association.

Boehm, C., A. Flaaen, and N. Pandalai-Nayar (2015). Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 TThoku Earthquake. *SSRN Electronic Journal*.

Boehm, J. and E. Oberfield (2020, November). Misallocation in the Market for Inputs: Enforcement and the Organization of Production*. *The Quarterly Journal of Economics 135*(4), 2007–2058.

Burstein, A., V. Carvalho, and B. Grassi (2020, October). Bottom-up Markup Fluctuations. Technical Report w27958, National Bureau of Economic Research, Cambridge, MA.

Burstein, A. and J. Cravino (2015, April). Measured Aggregate Gains from International Trade. *American Economic Journal: Macroeconomics 7*(2), 181–218.

Carvalho, C. (2006, December). Heterogeneity in Price Stickiness and the Real Effects of Monetary Shocks. *The B.E. Journal of Macroeconomics 6*(3), 1–58.

Carvalho, V. M. and B. Grassi (2019, April). Large Firm Dynamics and the Business Cycle. *American Economic Review 109*(4), 1375–1425.

Carvalho, V. M., M. Nirei, Y. Saito, and A. Tahbaz-Salehi (2016, December). Supply Chain Disruptions: Evidence from the Great East Japan Earthquake. SSRN Scholarly Paper ID 2893221, Social Science Research Network, Rochester, NY.

Carvalho, V. M. and A. Tahbaz-Salehi (2019). Production Networks: A Primer. *Annual Review of Economics 11*(1), 635–663. _eprint: https://doi.org/10.1146/annurev-economics-080218-030212.

Castro, N. (2019). The importance of production networks and sectoral heterogeneity for monetary policy. *Available at SSRN 3499902*.

Chaney, T. (2014, November). The Network Structure of International Trade. *American Economic Review 104*(11), 3600–3634.

Costinot, A. and A. Rodriguez-Clare (2014). Trade Theory with Numbers: Quantifying the Consequences of Globalization. In *Handbook of International Economics*, Volume 4, pp. 197–261. Elsevier.

Dew-Becker, I. (2021). Tail risk in production networks. Technical report, Working Paper.

Dew-Becker, I., A. Tahbaz-Salehi, and A. Vedolin (2021, December). Skewness and Time-Varying Second Moments in a Nonlinear Production Network: Theory and Evidence. SSRN Scholarly Paper 4026669, Social Science Research Network, Rochester, NY.

Dhyne, E., A. K. Kikkawa, and G. Magerman (2022, March). Imperfect Competition in Firm-to-Firm Trade. *Journal of the European Economic Association*, jvac013.

Dhyne, E., A. K. Kikkawa, M. Mogstad, and F. Tintelnot (2021, March). Trade and Domestic Production Networks. *The Review of Economic Studies 88*(2), 643–668.

di Giovanni, J., A. A. Levchenko, and I. Mejean (2018, January). The Micro Origins of International Business-Cycle Comovement. *American Economic Review 108*(1), 82–108.

Domar, E. D. (1961). On the Measurement of Technological Change. *The Economic Journal 71*(284), 709–729. Publisher: [Royal Economic Society, Wiley].

Edmond, C., V. Midrigan, and D. Y. Xu (2015, October). Competition, Markups, and the Gains from International Trade. *American Economic Review 105*(10), 3183–3221.

Edmond, C., V. Midrigan, and D. Y. Xu (2018, July). How Costly Are Markups? Technical

Report w24800, National Bureau of Economic Research, Cambridge, MA.

Elliott, M. and B. Golub (2022). Networks and economic fragility. *Annual Review of Economics 14*, 665–696.

Elliott, M., B. Golub, and M. V. Leduc (2022). Supply network formation and fragility. *American Economic Review 112*(8), 2701–47.

Flynn, J. P., C. Patterson, and J. Sturm (2022). Fiscal policy in a networked economy. Technical report, National Bureau of Economic Research.

Foerster, A., A. Hornstein, P.-D. Sarte, and M. Watson (2019). Aggregate implications of changing sectoral trends.

Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011, February). Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production. *Journal of Political Economy 119*(1), 1–38. Publisher: The University of Chicago Press.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica 79*(3), 733–772. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA8769.

Gopinath, G., S. Kalemli-Ozcan, L. Karabarbounis, and C. Villegas-Sanchez (2017, November). Capital Allocation and Productivity in South Europe*. *The Quarterly Journal of Economics 132*(4), 1915–1967.

Grassi, B. (2017). IO in I-O: Competition and Volatility in Input-Output Networks. Technical Report 1637, Society for Economic Dynamics. Publication Title: 2017 Meeting Papers.

Guerrieri, V., G. Lorenzoni, L. Straub, and I. Werning (2020, April). Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages? Technical Report w26918, National Bureau of Economic Research, Cambridge, MA.

Hall, R. E. (1990). Invariance properties of Solow's productivity residual. In *Growth, Productivity, Unemployment: Essays to Celebrate Bob Solow's Birthday, Ch. 5*. MIT Press.

Harberger, A. C. (1964). The Measurement of Waste. *The American Economic Review 54*(3), 58–76. Publisher: American Economic Association.

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

Hulten, C. R. (1973). Divisia index numbers. *Econometrica: Journal of the Econometric Society*, 1017–1025.

Hulten, C. R. (1978). Growth Accounting with Intermediate Inputs. *The Review of Economic Studies 45*(3), 511–518. Publisher: [Oxford University Press, Review of Economic

Studies, Ltd.].

Huneeus, F. (2018). Production Network Dynamics and the Propagation of Shocks. *Working paper*.

Huo, Z., A. A. Levchenko, and N. Pandalai-Nayar (2019, June). International Comovement in the Global Production Network. Working Paper 25978, National Bureau of Economic Research. Series: Working Paper Series.

Jones, C. I. (2011, January). Misallocation, Economic Growth, and Input-Output Economics. Working Paper 16742, National Bureau of Economic Research. Series: Working Paper Series.

Kehoe, T. J. and K. J. Ruhl (2008). Are shocks to the terms of trade shocks to productivity? *Review of Economic Dynamics 11*(4), 804–819.

Kikkawa, A. K., G. Magerman, and E. Dhyne (2019). Imperfect competition in firm-to-firm trade. *Available at SSRN 3389836*.

Kleinman, B., E. Liu, and S. J. Redding (2021, July). Dynamic Spatial General Equilibrium. Working Paper 29101, National Bureau of Economic Research. Series: Working Paper Series.

Kopytov, A., K. Nimark, B. Mishra, and M. Taschereau-Dumouchel (2022). Endogenous production networks under supply chain uncertainty.

La'O, J. and S. Bigio (2020, November). Distortions in Production Networks. *Quarterly Journal of Economics Volume 135(4)*, 2187–2253.

La'O, J. and A. Tahbaz-Salehi (2019). Optimal Monetary Policy in Production Networks. *SSRN Electronic Journal*.

Levinsohn, J. and A. Petrin (2003, April). Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies 70*(2), 317–341.

Lim, K. (2017). Endogenous production networks and the business cycle. *Working paper*.

Liu, E. (2019). Industrial policies and economic development. Technical report.

Liu, E. and A. Tsyvinski (2020, December). Dynamical Structure and Spectral Properties of Input-Output Networks. Technical Report w28178, National Bureau of Economic Research, Cambridge, MA.

Long, J. B. and C. I. Plosser (1983). Real Business Cycles. *Journal of Political Economy 91*(1), 39–69. Publisher: University of Chicago Press.

Magerman, G., K. De Bruyne, E. Dhyne, and J. Van Hove (2016, October). Heterogeneous firms and the micro origins of aggregate fluctuations. Working Paper Research 312, National Bank of Belgium.

McKenzie, L. (1951, December). Ideal output and the interdependence of firms. *The Economic Journal 61*(244), 785–803.

Meier, M. and T. Reinelt (2020, March). Monetary Policy,Markup Dispersion, and Aggregate TFP. Technical Report crctr224_2020_161, University of Bonn and University of Mannheim, Germany. Publication Title: CRC TR 224 Discussion Paper Series.

Nakamura, E. and J. Steinsson (2010, August). Monetary Non-Neutrality in a Multisector Menu Cost Model. *Quarterly Journal of Economics 125*(3), 961–1013.

Oberfield, E. (2018). A Theory of Input-Output Architecture. *Econometrica 86*(2), 559–589.

Osotimehin, S. and L. Popov (2020, February). Misallocation and Intersectoral Linkages. preprint, Institute Working Paper.

Pasten, E., R. Schoenle, and M. Weber (2017, August). Price Rigidity and the Origins of Aggregate Fluctuations. Working Paper 23750, National Bureau of Economic Research. Series: Working Paper Series.

Pasten, E., R. Schoenle, and M. Weber (2018, November). The Propagation of Monetary Policy Shocks in a Heterogeneous Production Economy. Working Paper 25303, National Bureau of Economic Research. Series: Working Paper Series.

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

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

Restuccia, D. and R. Rogerson (2017, August). The Causes and Costs of Misallocation. *Journal of Economic Perspectives 31*(3), 151–174.

Rubbo, E. (2020). Networks, Phillips Curves, and Monetary Policy. *Working paper*.

Rubbo, E. (2022). Monetary non-neutrality in the cross-section. Technical report.

Solow, R. M. (1957). Technical Change and the Aggregate Production Function. *The Review of Economics and Statistics 39*(3), 312–320. Publisher: The MIT Press.

Taschereau-Dumouchel, M. (2020). Cascades and fluctuations in an economy with an endogenous production network. *Available at SSRN 3115854*.

Vollrath, D. (2021). The elasticity of aggregate output with respect to capital and labor. *Available at SSRN 3835411*.

vom Lehn, C. and T. Winberry (2021, December). The Investment Network, Sectoral Comovement, and the Changing U.S. Business Cycle. *The Quarterly Journal of Economics 137*(1), 387–433.

A Proofs

Proof of Theorem 1. By definition,

$$d \log Y = \sum_i \frac{p_i y_i}{E} d \log y_i - \sum_i \frac{p_i l_i}{E} d \log l_i$$

where we define $y_i d \log y_i = 0$ if $y_i = 0$. Next observe that, by Assumption 2, we have

$$d \log x_i = \frac{\partial \log f_i}{\partial \log A_i} d \log A_i + \sum_j \frac{\partial \log f_i}{\partial \log x_{ij}} d \log x_{ij} + \sum_f \frac{\partial \log f_i}{\partial \log l_{if}} d \log x_{if}$$

and

$$y_i d \log y_i = x_i d \log x_i - \sum_j x_{ji} d \log x_{ji}.$$

If $\mu_i = 1$, then at the initial equilibrium, we have

$$\frac{\partial \log f_i}{\partial \log x_{ij}} = \frac{p_j x_{ij}}{p_i x_i}, \quad \frac{\partial \log f_i}{\partial \log l_{ij}} = \frac{p_j l_{ij}}{p_i x_i}$$

from Assumption 1. Hence

$$d \log y_i = \frac{x_i}{y_i} \left[d \log A_i + \sum_j \frac{\partial \log f_i}{\partial \log x_{ij}} d \log x_{ij} + \sum_h \frac{\partial \log f_i}{\partial \log l_{ih}} d \log l_{ih} \right] - \sum_j \frac{x_{ji}}{y_i} d \log x_{ji}.$$

In other words,

$$\frac{p_i y_i}{E} d \log y_i = \left[\frac{p_i x_i}{E} d \log A_i + \sum_j \frac{p_j x_{ij}}{E} d \log x_{ij} + \sum_h \frac{p_h l_{ih}}{E} d \log l_{ih} \right] - \sum_j \frac{p_i x_{ji}}{E} d \log x_{ji}$$

Putting this together yields

$$\begin{aligned} d \log Y &= \left[\sum_i \frac{p_i x_i}{E} d \log A_i + \sum_i \sum_j \frac{p_j x_{ij}}{E} d \log x_{ij} \right] - \sum_i \sum_j \frac{p_i x_{ji}}{E} d \log x_{ji} \\ &= \sum_i \frac{p_i x_i}{E} d \log A_i. \end{aligned}$$

as needed. Note that we did not need to fully specify how the allocation rule changed, we only imposed feasibility requirements imposed by Assumption 2. Hence, regardless of how the allocation rule changes, as long as it is feasible, the change in $d \log Y$ is given by Theorem 1.

□

Proof of Proposition 4. We have:

$$d \log y_i = \frac{x_i}{y_i} d \log x_i - \sum_{j \in \mathcal{N}} \frac{x_{ji}}{y_i} d \log x_{ji},$$

$$\mu_j^{-1}(d \log x_j - d \log A_j - \sum_{f \in \mathcal{H}} \frac{p_f l_{jf}}{p_j x_j} \mu_j d \log x_{jf}) = \sum_{i \in \mathcal{N}} \frac{p_i x_{ji}}{p_j x_j} d \log x_{ji}.$$

This equation is a consequence of cost minimization. We then have

$$\begin{aligned}
d \log Y &= \sum_{i \in \mathcal{N}} \frac{p_i y_i}{E} d \log y_i - \sum_{f \in \mathcal{H}} \frac{p_f l_f}{E} d \log l_f, \\
&= \sum_{i \in \mathcal{N}} \frac{p_i x_i}{E} d \log x_i - \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} \frac{p_i x_{ji}}{E} d \log x_{ji} - \sum_{f \in \mathcal{H}} \frac{p_f l_f}{E} d \log l_f, \\
&= \sum_{i \in \mathcal{N}} \frac{p_i x_i}{E} d \log x_i - \sum_{j \in \mathcal{I}} \frac{p_j x_j}{E} \mu_j^{-1}(d \log x_j - d \log A_j - \sum_{f \in \mathcal{H}} \frac{p_f y_{jf}}{p_j l_j} \mu_j d \log l_{jf}), \\
&= \sum_{i \in \mathcal{N}} \lambda_i \mu_i^{-1} d \log A_i + \sum_{i \in \mathcal{N}} \lambda_i (1 - \mu_i^{-1}) d \log y_i,
\end{aligned}$$

where the last line imposes that $d \log l_f = 0$ for every $f \in \mathcal{H}$. This last equation is an exact differential, and by differentiating it again, and evaluating the result at $\mu_i = 1$ for every i, we can derive the Hessian as needed.

Hence,

$$\begin{aligned}
\frac{\partial \log Y}{\partial \log \mu_k} &= \sum_{i \in \mathcal{N}} \lambda_i (1 - \mu_i^{-1}) \frac{\partial \log y_i}{\partial \log \mu_k} \\
\frac{\partial \log Y}{\partial \log A_j d \log \mu_k} &= \sum_{i \in \mathcal{N}} \frac{\partial \lambda_i}{\partial \log A_j} (1 - \mu_i^{-1}) \frac{\partial \log y_i}{\partial \log \mu_k} + \sum_{i \in \mathcal{I}} \lambda_i (1 - \mu_i^{-1}) \frac{\partial^2 \log y_i}{\partial \log A_j \partial \log \mu_k} \\
\frac{\partial \log Y}{\partial \log A_j d \log \mu_k} &= 0,
\end{aligned}$$

where the last line evaluates at $\mu_i = 1$ for every i. Since we assume that $\Delta \log Y$ is path-independent, this means that cross-partials are symmetric and hence all cross-partials between productivity and wedge shocks are zero at $\boldsymbol{\mu} = \mathbf{1}$.

It remains to characterize the Hessian with respect to only productivity shocks and changes in wedges separately. Again, from above, we know that

$$\frac{\partial \log Y}{\partial \log A_k} = \lambda_k,$$

whence

$$\frac{\partial^2 \log Y}{\partial \log A_j \partial \log A_k} = \frac{\partial \lambda_k}{\partial \log A_j}.$$

From above, we also know that

$$\frac{\partial \log Y}{\partial \log \mu_k} = \sum_{i \in \mathcal{N}} \lambda_i (1 - \mu_i^{-1}) \frac{\partial \log y_i}{\partial \log \mu_k}.$$

Hence,

$$\frac{\partial^2 \log Y}{\partial \log \mu_j \partial \log \mu_k} = \lambda_i \frac{\partial \log y_i}{\partial \log \mu_k}.$$

Putting all this together yields our desired result. □

B Generalization of Proposition 3

Following Baqaee and Farhi (2020c), we generalize the aggregation result in Proposition 3 to allow for multiple households, multiple factors, and non-constant-returns to scale. We partition the set of goods $\mathcal{N}$ in two. Without loss of generality, any non-constant-returns-to-scale production function can be represented as a composition of a constant-returns-to-scale production and a zero returns-to-scale endowment. That is, for any good $i \in \mathcal{N}$, we can write

$$x_i = f_i \left(\{x_{ij}\}_{j \in \mathcal{N}}, \{l_{ij}\}_{j \in \mathcal{H}}; A_i \right) = z_i \hat{f}_i(\{\frac{x_{ij}}{z_i}\}_{j \in \mathcal{N}}, \{\frac{l_{ij}}{z_i}\}_{j \in \mathcal{H}}; A_i),$$

with the requirement that $z_i = 1$. Written this way, x_i is produced with constant returns to scale once z_i is included as an input, and z_i is produced with zero returns to scale. Denote the set of all zero returns-to-scale endowments be denoted by $\mathcal{F}$.

To keep the notation simpler, we introduce the following assumption.

Assumption 3 (Non-negativity)**.** *For every $i \in \mathcal{N}$, assume that $y_i \geq 0$.*

This assumption rules out open economies.[26]

Once we impose Assumption 1, 2, and 3, we can generalize Theorem 1 to environments with inefficiencies in the following way.

Proposition 6 (Aggregation with Inefficiencies)**.** *Given changes in technology $d\log A$, wedges $d\log\mu$, and external input quantities $d\log l$, the change in aggregate real activity is given by*

$$d\log Y = \underbrace{\sum_{f\in\mathcal{H}}(\tilde{\Lambda}_f - \Lambda_f)d\log l_f}_{\text{external inputs}} + \underbrace{\sum_{i\in\mathcal{N}}\tilde{\lambda}_i d\log A_i}_{\text{technology}} \underbrace{- \sum_{i\in\mathcal{N}}\tilde{\lambda}_i d\log\mu_i - \sum_{f\in\mathcal{F}}\tilde{\Lambda}_f d\log\Lambda_f}_{\text{reallocation}}, \quad (28)$$

where $\tilde{\lambda}_i = \tilde{\Psi}_{Ci}$ and $\tilde{\Lambda}_f = \tilde{\Psi}_{Cf}$ denote cost-based Domar weights.

Proposition 6 shows that aggregation in inefficient equilibria requires knowledge of changes in factor income shares $d\log\Lambda$. These factor income shares, which are not needed when the equilibrium is efficient, matter because they encode how reallocations affect aggregate outcomes. To determine how these factor shares respond to primitive shocks to technologies and wedges, we need to use a generalized version of the forward and backward equations that allows for multiple factors and households (see Baqaee and Farhi, 2018). When there is only one primary factor, Proposition 6 collapses to Proposition 3.

[26] See Baqaee and Farhi (2019a) to see how this assumption can be relaxed. The primary reason to impose Assumption 3 is to simplify the notation.