



Annual Review of Economics

Production Networks: A Primer

Vasco M. Carvalho^{1,2,3} and Alireza Tahbaz-Salehi⁴

¹Faculty of Economics, University of Cambridge, Cambridge CB3 9DD, United Kingdom; email: vmpmde2@cam.ac.uk

²The Alan Turing Institute, London NW1 2DB, United Kingdom

³Centre for Economic Policy Research, London EC1V 0DX, United Kingdom

⁴Kellogg School of Management, Northwestern University, Evanston, Illinois 60208, USA; email: alirezat@kellogg.northwestern.edu

Annu. Rev. Econ. 2019. 11:635–63

The *Annual Review of Economics* is online at economics.annualreviews.org

<https://doi.org/10.1146/annurev-economics-080218-030212>

Copyright © 2019 by Annual Reviews.
All rights reserved

Keywords

networks, input–output linkages, shock propagation, aggregate fluctuations

Abstract

This article reviews the literature on production networks in macroeconomics. It presents the theoretical foundations for the role of input–output linkages as a shock propagation channel and as a mechanism for transforming microeconomic shocks into macroeconomic fluctuations. This article also provides a brief guide to the growing literature that explores these themes empirically and quantitatively.



1. INTRODUCTION

The idea that the production of goods and services relies on a complex web of transactions among a wide range of suppliers and customers has a long tradition in economics. As far back as the 1940s, in his study of the structure of the US economy, Wassily Leontief (1941, p. 3) observed that “layman and professional economist alike, practical planner and the subjects of his regulative activities, all are equally aware of the existence of some kind of interconnection between even the remotest parts of a national economy.” In this networked view of the production process, disturbances at certain firms or industries may spill over to other parts of the economy over input linkages, possibly transforming isolated microeconomic shocks into macroeconomic fluctuations. Indeed, Leontief (1941, p. 3) further argued that “the presence of these invisible but nevertheless very real ties can be observed whenever expanded automobile sales in New York City increase the demand for groceries in Detroit...when the sudden shutdown of the Pennsylvania coal mines paralyzes the textile mills in New England, and it reasserts itself with relentless regularity in alternative ups and downs of business cycles.”¹

In this article, we review the recent theoretical and empirical literature that revisits the role of production networks in propagating shocks and transforming microeconomic shocks into macroeconomic fluctuations. Although its motivation is classical, the modern literature on production networks relies on two fairly recent developments.

First, by drawing on tools from a diverse body of knowledge, the burgeoning field of network analysis has developed a conceptual framework and an extensive set of tools to effectively encode and measure interconnections among the units of analysis comprising a network. As we discuss below, when coupled with the language of general equilibrium theory, these serve as useful tools for assessing how shocks propagate throughout the economy, how different sectors comove over the business cycle, or how aggregate fluctuations can be traced out to localized microdisturbances.

Second, in tandem with developments elsewhere in economics, the availability of novel data sets on the granular nature of production has paved the way for a wide range of empirical and quantitative analyses seeking to answer classical questions, such as what the origins of aggregate fluctuations are. Almost 80 years after Leontief’s pioneering study of the structure of the US economy, modern input–output tables detail the complex patterns of input linkages across hundreds of industries. More significantly, it is now possible to identify the supplier–customer relations of millions of firms throughout the economy.

We provide a broad overview of the growing literature that leverages the above developments, with a particular focus on macroeconomic implications. While not a comprehensive survey, this article aims to offer a user guide to some of the recent theoretical and empirical work in the area. We begin, in Section 2, by presenting a benchmark model of production networks that serves as the basis for the main theoretical results in this article. We use this framework to demonstrate the role of input–output linkages as a shock propagation channel throughout the economy. In

¹It is worth noting that the emphasis on production networks goes well beyond the input–output literature originating in Leontief’s work and can be traced back to classical economics. By the mid-twentieth century, the development of general equilibrium theory also stressed such interdependencies. Indeed, von Neumann’s (1945) article “A model of general economic equilibrium” provides “a solution of a typical economic system” with the property that “goods are produced not only from ‘natural factors of production’ but in the first place from each other.” Such emphasis is also shared by the literature on multisector models of optimal growth. In particular, in an important precursor to the modern literature on production networks, Benhabib & Nishimura (1979) show that standard growth models featuring intersectoral linkages can exhibit optimal limit cycle trajectories, even in the absence of exogenous shocks. Given these antecedents, it is perhaps not surprising that the seminal contribution of Long & Plosser (1983) to modern business cycle theory—which we take as a starting point to our analysis—relies on a multisector general equilibrium foundation.

Section 3, we focus our attention on the role of input–output linkages as a mechanism for translating microeconomic shocks into aggregate fluctuations. These results provide sharp conditions for whether and when macroeconomic fluctuations can have their origins in idiosyncratic shocks to individual firms or disaggregated industries. In Section 4, we review the growing literature that empirically and quantitatively explores the role of production networks at the firm and industry levels. We conclude in Section 5 by discussing several open questions and promising avenues for future research.

2. A MODEL OF PRODUCTION NETWORKS

We start by presenting a baseline model of production networks, which serves as a useful starting point for analysis. The model is a static variant of the multisector general equilibrium model of Long & Plosser (1983), which is also analyzed by Acemoglu et al. (2012). We discuss various modifications and generalizations of this model in the subsequent sections.

2.1. Baseline Model

Consider a static economy consisting of n competitive industries denoted by $\{1, 2, \dots, n\}$, each producing a distinct product. Each product can be either consumed by the households or used as an intermediate input for production of other goods. Firms in each industry employ Cobb–Douglas production technologies with constant returns to scale to transform intermediate inputs and labor into final products. In particular, the output of industry i is given by

$$y_i = z_i \zeta_i l_i^{\alpha_i} \prod_{j=1}^n x_{ij}^{a_{ij}}, \quad 1.$$

where l_i is the amount of labor hired by firms in industry i , x_{ij} is the quantity of good j used for production of good i , $\alpha_i > 0$ denotes the share of labor in industry i 's production technology, z_i is a Hicks-neutral productivity shock, and ζ_i is some normalization constant the value of which only depends on model parameters.²

The exponents $a_{ij} \geq 0$ in Equation 1 formalize the idea that firms in an industry may need to rely on the goods produced by other industries as intermediate inputs for production. In particular, a larger a_{ij} means that good j is a more important input for the production of good i , whereas $a_{ij} = 0$ means that good j is not a necessary input for i 's production. Note that, in general, $a_{ij} \neq a_{ji}$, as industry i 's reliance on industry j as an input supplier may be different from j 's dependence on i . Furthermore, it may also be the case that $a_{ii} > 0$, as good i may itself be used as an intermediate input for production by firms in industry i . Finally, note that the assumption that all technologies exhibit constant returns to scale implies that $\alpha_i + \sum_{j=1}^n a_{ij} = 1$ for all i .

In addition to the firms described above, the economy is populated by a representative household, which supplies one unit of labor inelastically and has logarithmic preferences over the n goods given by

$$u(c_1, \dots, c_n) = \sum_{i=1}^n \beta_i \log(c_i / \beta_i), \quad 2.$$

²Below, we set the value of this constant to $\zeta_i = \alpha_i^{-\alpha_i} \prod_{j=1}^n a_{ij}^{-a_{ij}}$. This choice has no bearing on the results, as the sole purpose of this constant is to simplify the analytical expressions.

where c_i is the amount of good i consumed. The constants $\beta_i \geq 0$ measure various goods' shares in the household's utility function, normalized such that $\sum_{i=1}^n \beta_i = 1$.

Equations 1 and 2 thus fully specify the environment. The competitive equilibrium of this economy is defined in the usual way: It consists of a collection of prices and quantities such that (a) the representative household maximizes its utility, (b) the representative firm in each sector maximizes its profits while taking the prices and the wage as given, and (c) all markets clear.

Before characterizing the equilibrium, it is useful to define a few key concepts that will play a central role in our subsequent analysis. First, note that we can summarize the input–output linkages between various industries with a matrix $\mathbf{A} = [a_{ij}]$, which, with some abuse of terminology, we refer to as the economy's input–output matrix.³ Thus, coupled with the vector of productivity shocks (z_1, \dots, z_n) , the input–output matrix \mathbf{A} serves as a sufficient statistic for the production side of the economy. Note that the assumption that $\alpha_i > 0$ for all i implies that \mathbf{A} is an element-wise nonnegative matrix with row sums that are strictly less than 1. This in turn guarantees that the spectral radius of \mathbf{A} —defined as the largest absolute value of its eigenvalues—is strictly less than 1 (Berman & Plemmons 1979, p. 37).

The input–output linkages between various industries can alternatively be represented by a weighted directed graph on n vertices. Each vertex in this graph—which we refer to as the economy's production network—corresponds to an industry, with a directed edge with weight $a_{ij} > 0$ present from vertex j to vertex i if industry j is an input supplier of industry i . While the production network representation of the economy is equivalent to the representation using the input–output matrix,⁴ it can provide a conceptually simpler framework for summarizing (and visualizing) input–output linkages.

Finally, we define an industry's Domar weight as that industry's sales as a fraction of GDP. More specifically, the Domar weight of industry i is defined as

$$\lambda_i = \frac{p_i y_i}{\text{GDP}}, \quad 3.$$

where p_i is the price of good i and y_i is industry i 's output. These weights play a key role in the analysis in Section 3.

We now proceed to determining the equilibrium prices and quantities. First, note that firms in industry i choose their demand for labor and intermediate goods to maximize profits, $\pi_i = p_i y_i - w l_i - \sum_{j=1}^n p_j x_{ij}$, while taking all prices (p_1, \dots, p_n) and the wage w as given. Thus, the first-order conditions corresponding to firms in industry i are given by $x_{ij} = a_{ij} p_i y_i / p_j$ and $l_i = \alpha_i p_i y_i / w$. Plugging these expressions into firm i 's production function in Equation 1 and taking logarithms imply that

$$\log(p_i/w) = \sum_{j=1}^n a_{ij} \log(p_j/w) - \epsilon_i,$$

where $\epsilon_i = \log z_i$ is the (log) productivity shock to firms in industry i . Since the above relationship has to hold for all industries i , it provides a system of equations to solve for all relative prices in terms of productivity shocks. More specifically, rewriting this system of equations in matrix form implies that $\hat{p} = \mathbf{A}\hat{p} - \epsilon$, where \mathbf{A} is the economy's input–output matrix and

³More generally, the input–output matrix $\mathbf{\Omega} = [\omega_{ij}]$ of an economy is defined in terms of input expenditures as a fraction of sales, that is, $\omega_{ij} = p_j x_{ij} / p_i y_i$. However, in the special case that all technologies and preferences are Cobb–Douglas, ω_{ij} coincides with the exponent a_{ij} in Equation 1.

⁴In fact, in graph theory terminology, the input–output matrix \mathbf{A} is nothing but the (weighted) adjacency matrix of the economy's production network.

$\hat{p} = (\log(p_1/w), \dots, \log(p_n/w))'$ and $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ denote the vectors of log relative prices and productivity shocks, respectively. Consequently, the equilibrium vector of (log) relative prices is given by

$$\hat{p} = -(\mathbf{I} - \mathbf{A})^{-1}\epsilon \quad 4.$$

in terms of industry-level shocks and the economy's production network.

Before proceeding any further, it is useful to comment on some of the key properties of matrix $\mathbf{L} = (\mathbf{I} - \mathbf{A})^{-1}$ in Equation 4, commonly known as the economy's Leontief inverse. First, the fact that the input-output matrix \mathbf{A} is nonnegative with a spectral radius that is strictly less than 1 means that $\mathbf{I} - \mathbf{A}$ is a nonsingular M-matrix, which in turn guarantees that the Leontief inverse $\mathbf{L} = (\mathbf{I} - \mathbf{A})^{-1}$ always exists and is element-wise nonnegative.⁵ Second, the observation that the spectral radius of \mathbf{A} is strictly less than 1 also implies that the Leontief inverse can be expressed as the infinite sum of the powers of the input-output matrix \mathbf{A} (Stewart 1998, p. 55), i.e.,

$$\mathbf{L} = (\mathbf{I} - \mathbf{A})^{-1} = \sum_{k=0}^{\infty} \mathbf{A}^k. \quad 5.$$

This decomposition illustrates that the (i, j) element of the Leontief inverse measures the importance of industry j as a direct and indirect input supplier to industry i in the economy. To see this, note that, for any $i \neq j$, Equation 5 implies that $\ell_{ij} = a_{ij} + \sum_{r=1}^n a_{ir}a_{rj} + \dots$, with the first term in this expression accounting for j 's role as a direct supplier to i , the second term accounting for j 's role as a supplier to i 's suppliers, and so on. Interpreted in terms of the production network representation of the economy, ℓ_{ij} accounts for all possible directed walks (of various lengths) that connect industry j to industry i over the network.

Returning to equilibrium characterization, recall that the firms' first-order conditions imply that the quantity demanded by industry i from industry j is given by $x_{ij} = a_{ij}p_i y_i / p_j$, while the representative household's logarithmic utility implies that it demands $c_j = \beta_j w / p_j$ units of good j . Plugging these expressions into the market-clearing condition for good j , which is given by $y_j = c_j + \sum_{i=1}^n x_{ij}$, implies that $p_j y_j = \beta_j w + \sum_{i=1}^n a_{ij} p_i y_i$. Dividing both sides of this equation by w and noting that the value added in this economy is equal to the household's labor income, we obtain

$$\lambda_j = \beta_j + \sum_{i=1}^n a_{ij} \lambda_i,$$

where λ_i is the Domar weight of industry i defined in Equation 3. Rewriting the above equation in matrix form and solving for the vector of Domar weights implies that $\lambda = (\mathbf{I} - \mathbf{A}')^{-1} \beta$ or, equivalently, $\lambda_i = p_i y_i / \text{GDP} = \sum_{j=1}^n \beta_j \ell_{ji}$. Furthermore, recall from Equation 4 that $\log(p_i / \text{GDP}) = -\sum_{j=1}^n \ell_{ij} \epsilon_j$, thus leading to the following result:

Theorem 1. The log output of industry i is given by

$$\log(y_i) = \sum_{j=1}^n \ell_{ij} \epsilon_j + \delta_i, \quad 6.$$

where δ_i is some constant that is independent of the shocks.

⁵ A square matrix \mathbf{Q} is called an M-matrix if there exist a nonnegative square matrix \mathbf{B} and a constant $r \geq \rho(\mathbf{B})$ such that $\mathbf{Q} = r\mathbf{I} - \mathbf{B}$, where $\rho(\mathbf{B})$ is the spectral radius of \mathbf{B} . If $r > \rho(\mathbf{B})$, then \mathbf{Q} is a nonsingular M-matrix. Plemmons (1977, theorem 2) shows that the inverse of any nonsingular M-matrix is element-wise nonnegative.

The above theorem has a few important implications. First, the mere fact that the output of industry i may depend on the shocks to industries $j \neq i$ indicates that the input–output linkages in the economy can function as a mechanism for the propagation of shocks from one industry to another. Second, it shows that the resulting propagation patterns are captured by the economy's Leontief inverse \mathbf{L} (and not its input–output matrix \mathbf{A}). This means that the input–output linkages can result in both direct and indirect propagation of the shocks over the production network. Third, the fact that the impact of a shock to industry j on i 's output is captured by ℓ_{ij} means that productivity shocks in this model propagate downstream from one industry to its customers, its customers' customers, and so on.⁶ To see this, recall from the expansion in Equation 5 that ℓ_{ij} is a measure of the importance of industry j as (direct and indirect) input supplier to industry i .⁷

The intuition underlying Theorem 1 is fairly straightforward. Suppose that industry j is hit by a negative shock that reduces its production and thus increases the price of good j . Such a price increase adversely impacts all of the industries that rely on good j as an intermediate input for production, thus creating a direct impact on j 's customer industries. However, this initial impact will in turn result in further propagation over the production network: The prices of goods produced by industries affected in the first round of propagation will rise, creating an indirect negative effect on their own customer industries, and so on. The overall effect of these direct and indirect downstream propagations of the initial shock is summarized by the corresponding element of the economy's Leontief inverse.

Why is it that productivity shocks in this model only propagate from an industry to its (direct and indirect) customers, and not to its suppliers? The absence of such upstream propagation is a consequence of three specific features of the model: (a) Cobb–Douglas preferences and technologies, (b) a single factor of production (in this case labor), and (c) constant returns to scale. The latter two features together guarantee that productivity shocks do not impact upstream prices (relative to the wage): The price of good i is equal to industry i 's marginal cost, which only depends on the productivities of i and its upstream industries. Furthermore, recall that, in a Cobb–Douglas economy, the Domar weights of all industries are invariant to the shocks ($\lambda_i = p_i y_i / \text{GDP} = \sum_{j=1}^n \beta_j \ell_{ji}$). Thus, the absence of upstream effects on (relative) prices translates into the absence of upstream effects on quantities.

We now turn to determining the production network's macroeconomic implications. Recall from Equation 4 that the log relative price of good i is given by $\log(p_i/w) = -\sum_{j=1}^n \ell_{ij} \epsilon_j$. Multiplying both sides by β_i and summing over all industries i leads to $\log(\text{GDP}) = \sum_{i,j=1}^n \beta_i \ell_{ij} \epsilon_j + \sum_{i=1}^n \beta_i \log p_i$. Choosing the consumption good bundle, the price of which is given by $P_c = \prod_{i=1}^n p_i^{\beta_i}$, as the numeraire implies that $\sum_{i=1}^n \beta_i \log p_i = 0$. We therefore have the following result:

Theorem 2. The economy's (log) real value added is given by

$$\log(\text{GDP}) = \sum_{i=1}^n \lambda_i \epsilon_i, \quad 7.$$

where

$$\lambda_i = \frac{p_i y_i}{\text{GDP}} = \sum_{j=1}^n \beta_j \ell_{ji}, \quad 8.$$

and ℓ_{ji} is the (j, i) element of the economy's Leontief inverse $\mathbf{L} = (\mathbf{I} - \mathbf{A})^{-1}$.

⁶As we show in the sections below, demand-side shocks exhibit significantly different propagation patterns.

⁷Note that, when the production network exhibits cycles (say, in an economy with roundabout production), an industry can be simultaneously upstream and downstream of another industry. What we mean by downstream propagation is that shocks transmit from one industry to another in the direction of the flow of goods and services.

The significance of the above result is twofold. First, Equation 7 illustrates that (log) aggregate output is a linear combination of industry-level productivity shocks with coefficients that are given by the industries' Domar weights. Thus, the Domar weight of industry i is a sufficient statistic for how shocks to that industry impact aggregate output. As we discuss in Section 2.3.2, some variant of this relationship, which is commonly known as Hulten's theorem, holds much more generally (Hulten 1978, Gabaix 2011).

Second, Theorem 2 establishes that, with Cobb-Douglas preferences and technologies, the Domar weights take a particularly simple form: The Domar weight of each industry depends only on the preference shares and the corresponding column of the economy's Leontief inverse. This means that, while λ_i is a sufficient statistic for how shocks to industry i impact $\log(\text{GDP})$, the value of λ_i itself depends on the economy's production network. In particular, as Equation 8 illustrates, all else equal, an increase in ℓ_{ji} will increase industry i 's Domar weight and thus intensify the impact of shocks to i on aggregate output. The intuition underlying this result parallels that of Theorem 1: The downstream propagation of shocks from an industry to its direct and indirect customers means that, all else equal, shocks to industries that are more important input suppliers to the rest of the economy have a more pronounced effect on macroeconomic aggregates.

2.2. Demand-Side Shocks

In this section, we show that demand-side shocks lead to propagation patterns that are substantially different from those of supply-side productivity shocks discussed above.

To incorporate demand-side shocks into the model, we follow Acemoglu et al. (2016a) and modify our benchmark model by assuming that the government purchases an exogenously given quantity g_i of good i . This modification implies that good i 's market-clearing condition is given by $y_i = c_i + g_i + \sum_{j=1}^n x_{ji}$. Thus, changes in government spending on various goods correspond to demand-side shocks that affect industries differentially. To simplify the derivations, we abstract away from supply-side shocks by assuming that $z_i = 1$ for all i .

Solving for the economy's competitive equilibrium is straightforward. Plug industry i 's first-order conditions—given by $x_{ij} = a_{ij}p_j y_i / p_j$ and $l_i = \alpha_i p_i y_i / p_j$ —into Equation 1 and solve the resulting system of equations, which implies that $p_i = w$ for all i . This means that, unlike productivity shocks, demand-side shocks do not impact relative prices. Furthermore, the representative household's budget constraint is given by $\sum_{i=1}^n p_i c_i = w - T$, where $T = \sum_{i=1}^n p_i g_i$ is the total amount of government spending, financed by lump sum taxes on the household. Therefore, the market-clearing condition for good i reduces to $y_i = \beta_i(1 - \sum_{j=1}^n g_j) + g_i + \sum_{j=1}^n a_{ji} y_j$. Rewriting the resulting system of equations in matrix form, we obtain $y = (1 - g' \mathbf{1})\beta + g + \mathbf{A}'y$, where $g = (g_1, \dots, g_n)'$ is the vector of quantities demanded by the government and $\mathbf{1}$ is a vector with all entries equal to 1. Solving this system of equations leads to the following result:

Theorem 3. The output of industry i is given by

$$y_i = \sum_{j=1}^n \ell_{ji} g_j + \left(1 - \sum_{k=1}^n g_k\right) \left(\sum_{j=1}^n \ell_{ji} \beta_j\right), \quad 9.$$

where $\mathbf{L} = (\mathbf{I} - \mathbf{A})^{-1}$ is the economy's Leontief inverse matrix.

Contrasting Theorems 1 and 3 illustrates the stark difference in how supply- and demand-side shocks propagate: Whereas the impact of a productivity shock to industry j on the output of industry i is captured by ℓ_{ij} , the impact of a demand shock to j on i is captured via ℓ_{ji} . This

means that, unlike supply-side shocks that propagate downstream, demand-side shocks propagate upstream from one industry to its direct and indirect suppliers.⁸ The intuition underlying this propagation pattern is as follows: A positive demand shock to industry j increases j 's demands for inputs, which is in effect a positive demand shock to j 's suppliers. A similar logic implies that the original demand shock would propagate further upstream.

We conclude this discussion by noting that the baseline model that we focus on above is special along multiple dimensions: It is a perfectly competitive economy with a single factor of production and Cobb-Douglas technologies and preferences. In the remainder of this section, we briefly discuss the implications of relaxing some of these assumptions.

2.3. More General Production Technologies

We start by illustrating how relaxing the assumption that all production technologies are Cobb-Douglas alters shocks' propagation patterns as well as their aggregate implications.

2.3.1. Propagation patterns. One of the consequences of assuming Cobb-Douglas production technologies is that an industry's expenditure on various inputs as a fraction of its sales is invariant to the realization of the shocks. In particular, for any pair of industries i and j , the ratio $\omega_{ij} = p_j x_{ij} / p_i y_i$ is equal to the exponent a_{ij} in Equation 1, which is an exogenously given parameter of the model. Such an invariance, however, no longer holds for more general production technologies. This in turn can lead to richer propagation patterns over input-output linkages.

These effects are explored by Carvalho et al. (2016), who focus on a generalization of the baseline model by replacing the production functions in Equation 1 by a nested constant elasticity of substitution (CES) structure. Since a closed-form characterization is in general not possible, they use a first-order approximation to show that, when the elasticities of substitution between various intermediate inputs or between the intermediates and primary factors of production are different from 1, a negative productivity shock to industry i impacts the output of other industries via two distinct channels. First, the resulting increase in good i 's price adversely impacts all industries that rely on good i as an intermediate input for production, thus leading to a downstream propagation of the shock to i 's direct and indirect customers. This output effect thus leads to propagation patterns that mirror those in a Cobb-Douglas economy. Second, productivity shocks to industry i may also result in reallocation of resources across different industries depending on the elasticities of substitution across various inputs. For instance, the increase in the price of good i in response to a negative shock to i results in an increase (decrease) in demand by i 's customers for input $j \neq i$ if goods i and j are gross substitutes (complements) in these customers' production technologies. Thus, in contrast to the Cobb-Douglas economy, the impact of a productivity shock to industry i may not remain confined to i 's downstream industries.

These results are further extended by Baqaee & Farhi (2018b) to a general class of economies with heterogeneous agents, arbitrary nested CES production structures, and multiple (and potentially industry-specific) factors of production. For the purposes of this article, we find it instructive to focus on a special case with a single factor of production (labor) and a single CES nest to clarify the two propagation channels highlighted in the previous paragraph. In particular, suppose that

⁸Note that, in addition to the upstream propagation channel highlighted above, the expression in Equation 9 also includes a term $(1 - \sum_{k=1}^n g_k)(\sum_{j=1}^n \ell_{ji} \beta_j)$ that corresponds to a resource constraint effect: An increase in government spending requires higher taxes on the household and thus fewer resources for private consumption.

the production technology of firms in industry i is given by

$$y_i = z_i \zeta_i l_i^{\alpha_i} \left(\sum_{j=1}^n a_{ij}^{1/\sigma_i} x_{ij}^{1-1/\sigma_i} \right)^{(1-\alpha_i)\sigma_i/(\sigma_i-1)}, \quad 10.$$

where $\alpha_i + \sum_{j=1}^n a_{ij} = 1$, σ_i denotes the elasticity of substitution between the various inputs and the normalization constant $\zeta_i = \alpha_i^{-\alpha_i} (1 - \alpha_i)^{-(1-\alpha_i)\sigma_i/(\sigma_i-1)}$. This economy reduces to the baseline model with the Cobb-Douglas technologies in Equation 1 when $\sigma_i \rightarrow 1$ for all i . As we show in the **Supplemental Appendix**, log-linearization of equilibrium conditions implies that the effect of a shock to industry j on the output of industry i , up to a first-order approximation, is given by

$$\frac{d \log y_i}{d \log z_j} \Big|_{\log z=0} = \ell_{ij} + \frac{1}{\lambda_i} \sum_{k=1}^n (\sigma_k - 1) \lambda_k \underbrace{\left[\sum_{r=1}^n a_{kr} \ell_{ri} \ell_{rj} - \frac{1}{1 - \alpha_k} \left(\sum_{r=1}^n a_{kr} \ell_{ri} \right) \left(\sum_{r=1}^n a_{kr} \ell_{rj} \right) \right]}_{Q_k}. \quad 11.$$

As above, λ_i denotes industry i 's Domar weight and $\mathbf{L} = (\mathbf{I} - \mathbf{A})^{-1}$ is the economy's Leontief inverse, where $\mathbf{A} = [a_{ij}]$.

The first term on the right-hand side of Equation 11 is in line with the expression in Equation 6 and captures the downstream output effect that is also present in a Cobb-Douglas economy. The second term captures the reallocation effect: In response to a negative shock to industry j , all industries k that are downstream of j may readjust their demand for all other inputs. Crucially, the impact of such readjustments by any given k on the output of industry i depends on (a) the size of industry k as captured by its Domar weight λ_k , (b) the elasticity of substitution σ_k in k 's production function, and (c) the extent to which the supply chains that connect i and j to k coincide with one another.

To clarify the workings of the reallocation channel and its relationship to Equation 11, it is instructive to focus on a simple environment. Consider an industry k with $\sigma_k > 1$ that is downstream of both i and j , each of which supply k via a single production chain. This means that the constellation of these industries in the production network can take one of the following two forms: Either i and j supply industry k via production chains that pass through the same supplier of k , as depicted in **Figure 1a**, or the production chains supplied by i and j reach industry k via two distinct suppliers of k , as depicted in **Figure 1b**.

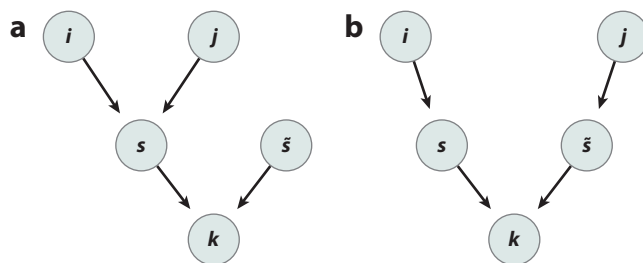


Figure 1

Two simple production networks where each vertex corresponds to an industry, with a directed edge present from one vertex to another if the former is an input supplier (s, \tilde{s}) to the latter. Industry k is downstream of industries i and j in either case. In panel *a*, the input paths from i and j to k overlap; in panel *b*, they do not.

It is not hard to verify that, for the economy depicted in **Figure 1a**, expression Q_k on the right-hand side of Equation 11 is equal to

$$Q_k = \sum_{r=1}^n a_{kr} \ell_{ri} \ell_{rj} - \frac{1}{1 - \alpha_k} \left(\sum_{r=1}^n a_{kr} \ell_{ri} \right) \left(\sum_{r=1}^n a_{kr} \ell_{rj} \right) = a_{ks} \ell_{si} \ell_{sj} \left(1 - \frac{a_{ks}}{1 - \alpha_k} \right),$$

where s is the supplier of k that is downstream to both i and j . This expression is strictly positive as long as industry s is not the sole supplier of k (i.e., $a_{ks} < 1 - \alpha_k$). Thus, a negative productivity shock to industry j results in a decrease in i 's output. This is, of course, fairly intuitive: The fact that $\sigma_k > 1$ implies that, in response to a negative shock to j , industry k substitutes away from the production chain supplied by j , in the process also reducing the demand for industry i 's output.

The propagation patterns induced by the substitution channel are different in the economy depicted in **Figure 1b**. Once again, a negative shock to j would force industry k to substitute away from the production chain that is supplied by j whenever $\sigma_k > 1$. However, unlike the previous case, such a substitution results in an increase in i 's output precisely because the production chains supplied by i and j do not overlap with one another. Indeed, expression Q_k on the right-hand side of Equation 11 is given by

$$Q_k = \sum_{r=1}^n a_{kr} \ell_{ri} \ell_{rj} - \frac{1}{1 - \alpha_k} \left(\sum_{r=1}^n a_{kr} \ell_{ri} \right) \left(\sum_{r=1}^n a_{kr} \ell_{rj} \right) = \frac{-1}{1 - \alpha_k} (a_{ks} \ell_{si}) (a_{k\bar{s}} \ell_{\bar{s}j}),$$

which is strictly negative, as expected.

These simple examples illustrate that, when inputs are gross substitutes, more overlap in the production chains that originate from i and j translates a negative productivity shock to j into a reduction in i 's output, while such overlaps have the opposite effect when inputs are gross complements.

2.3.2. Aggregate effects and Hulten's theorem. Recall from Theorem 2 that, with Cobb-Douglas preferences and technologies, an industry's Domar weight is a sufficient statistic for how total factor productivity (TFP) shocks to that industry impact GDP. In this section, we show that a variant of this relationship holds much more generally: In any efficient economy, the impact on output of a TFP shock to industry i is equal to i 's Domar weight up to a first-order approximation. More specifically, if z_i denotes the TFP shock to industry i , then irrespective of the household's preferences and the firms' production technologies,

$$\frac{d \log(\text{GDP})}{d \log(z_i)} = \lambda_i, \quad 12.$$

where $\lambda_i = p_i y_i / \text{GDP}$ is the Domar weight of industry i .

The simplicity of the relationship in Equation 12, which has come to be known as Hulten's theorem, makes it a useful tool in empirical studies of microeconomic origins of aggregate fluctuations. For instance, as we discuss in subsequent sections, Gabaix (2011) uses the empirical distribution of firm-level Domar weights to measure the extent to which firm-level shocks can explain GDP volatility, while Carvalho & Gabaix (2013) rely on Hulten's theorem to investigate whether

changes in the economy's microeconomic composition can account for the Great Moderation and its unraveling in major world economies.

Despite its simplicity, Hulten's theorem may appear surprising at first sight: How is it that, in the presence of input–output linkages, an industry's role in shaping aggregate outcomes is entirely reflected by its size, irrespective of its position in the production network?⁹

To derive and illustrate the intuition behind Equation 12, we follow papers such as those of Gabaix (2011) and Baqaee & Farhi (2018a) and extend the baseline model in Section 2 by allowing for general production functions, preferences, and factor markets. More specifically, consider a static economy consisting of n competitive industries, each producing a distinct product using intermediate inputs and m different primary factors of production. Firms in industry i employ constant returns to scale production technologies given by

$$y_i = z_i f_i(x_{i1}, \dots, x_{im}, l_{i1}, \dots, l_{im}),$$

where z_i is the Hicks-neutral productivity shock to industry i , and x_{ij} and l_{ik} are the quantities of good j and the k th primary factor used by firms in industry i , respectively. The economy is also populated by a representative household with preferences $u(c_1, \dots, c_n)$, which we assume to be homogenous of degree 1. This representative household is endowed with b_k units of the k th primary factor, which it supplies inelastically to the market. As above, we focus on the economy's competitive equilibrium, in which (a) all firms maximize their profits, taking the factor and intermediate good prices as given; (b) the representative household maximizes its utility; and (c) all good and factor markets clear.

By the first welfare theorem, the competitive equilibrium of this economy is efficient. This means that one can determine the equilibrium allocation by solving the social planner's problem:

$$\begin{aligned} W = \max_{c_i, l_{ik}, x_{ij}} \quad & u(c_1, \dots, c_n) \\ \text{s.t.} \quad & c_i + \sum_{j=1}^n x_{ji} = z_i f_i(x_{i1}, \dots, x_{im}, l_{i1}, \dots, l_{im}) \quad i = 1, \dots, n \\ & \sum_{i=1}^n l_{ik} = b_k \quad k = 1, \dots, m. \end{aligned}$$

The constraints in the above problem correspond to the resource constraints for good i and the k th primary factor of production, respectively. The first-order condition of optimality requires that $du/dc_i = \eta_i$, where η_i is the Lagrange multiplier corresponding to good i 's resource constraint. Furthermore, applying the envelope theorem to the planner's problem implies that $dW/dz_i = \eta_i f_i(x_{i1}, \dots, x_{im}, l_{i1}, \dots, l_{im}) = \eta_i y_i/z_i$. Consequently, we have

$$\frac{d \log(W)}{d \log(z_i)} = \frac{\eta_i y_i}{W}. \quad 13.$$

⁹The apparent discrepancy between one's intuition and Hulten's theorem is probably best captured in a speech by Summers (2013): "There would be a set of economists who would sit around explaining that electricity was only 4% of the economy, and so if you lost 80% of electricity, you couldn't possibly have lost more than 3% of the economy." However, "we would understand that...when there wasn't any electricity, there wasn't really going to be much economy."

Next, note that the household's optimization problem in the decentralized representation of the equilibrium is given by

$$\begin{aligned} W &= \max_{c_i} u(c_1, \dots, c_n) \\ \text{s.t.} \quad & \sum_{i=1}^n p_i c_i = \sum_{k=1}^m w_k b_k, \end{aligned}$$

where w_k is the price of the k th primary factor of production. First-order conditions imply that $du/dc_i = \phi p_i$, where ϕ is the Lagrange multiplier corresponding to the household's budget constraint. Contrasting this with the corresponding first-order condition from the planner's problem implies that $\eta_i = \phi p_i$. Furthermore, multiplying both sides of the household's first-order condition by c_i , summing over all i , and applying Euler's homogenous function theorem to the utility function u implies that $W = \phi \sum_{i=1}^n p_i c_i$. Replacing for η_i and W in Equation 13 then establishes Hulten's theorem in Equation 12.

The above derivations illustrate that equilibrium efficiency and the envelope theorem lie at the heart of Hulten's theorem. In general, a positive productivity shock to industry i impacts aggregate output via two channels. First, it results in an outward shift in the economy's production possibility frontier. Second, it may result in the reallocation of resources across the various industries in the economy. However, when the original allocation is efficient, any aggregate effect due to the resource reallocation channel is second order (by the envelope theorem) and thus can be ignored in a first-order approximation. This observation also implies that Hulten's theorem may not hold in inefficient economies. Indeed, Equation 12 is violated in the production network models of Jones (2013), Bigio & La'O (2017), and Liu (2018), all of which exhibit some form of distortions or wedges.

We also remark that, while Hulten's theorem establishes that Domar weights are sufficient statistics for how industry-level shocks impact aggregate output, these weights are endogenous objects that are determined in equilibrium. In fact, as Theorem 2 illustrates, even in the very simple economy of Section 2.1, with Cobb-Douglas preferences and technologies, Domar weights depend on the economy's production network (via its Leontief inverse) and the household's preferences.

Finally, it is important to bear in mind that Hulten's theorem maintains that Domar weights are sufficient statistics for microeconomic shocks' aggregate impact only up to a first order. This means that, while Equation 12 can be a reasonable approximation when either the shocks are small or the economy does not exhibit significant nonlinearities, it may be a fairly poor approximation more generally.¹⁰ Baqaee & Farhi (2018a) explore the role of such nonlinearities by extending Hulten's theorem to include the second-order effects of microeconomic shocks on aggregate output.¹¹ Focusing on a general economy with a nested CES structure, they illustrate that these second-order terms depend on the economy's production network, the elasticities of substitution at various CES nests, and the degree to which factors can be reallocated across industries. The

¹⁰ Put differently, Equation 12 is obtained under the assumption that the shocks' impact on the Domar weights themselves is negligible. In general, however, the sales share of an industry may respond significantly to shocks. This observation also illustrates why Hulten's theorem holds globally (i.e., regardless of the size of the shocks) in the baseline model of Section 2.1 (Equation 7): In the special case that all preferences and technologies are Cobb-Douglas, Domar weights are independent of the realization of productivity shocks, which implies that Hulten's first-order approximation is exact.

¹¹ Relatedly, Acemoglu et al. (2016b) study the micro and macro implications of nonlinearities in a reduced-form model of network interactions.

presence of these nonlinearities (which include, but are not restricted to, the second-order effects) is at the core of the apparent disparity between Hulten's theorem and one's intuition regarding network linkages mentioned above: While Hulten's theorem is a statement about the shocks' first-order effects, the economy's production network can manifest itself via significant nonlinear effects captured by the higher-order terms.

2.4. Frictions and Market Imperfections

As emphasized above, propagation of productivity shocks in the perfectly competitive models of Sections 2.1 and 2.3 occurs via two channels. First, a negative shock to industry i results in an increase in the price of good i , thus increasing the production cost of industries that use i as an input for production. Second, the increase in i 's price may also induce the customer industries to readjust their demand for other intermediate inputs. These observations imply that departures from the assumption of perfect competition that either (a) distort the input usage of customer industries or (b) modify prices' responsiveness to the shocks can reshape propagation patterns over the network.

The simplest departure from the assumption of perfect competition is the introduction of exogenous wedges—say, in the form of markups—between firms' marginal revenue and marginal costs that distort their input and output choices away from efficient levels. This is the approach adopted by Jones (2013), Bigio & La'O (2017), and Fadinger et al. (2018), who investigate how the production network interacts with productivities and markups in the determination of aggregate outcomes in the Cobb-Douglas economy of Section 2.1.¹² They find that the extent of resource misallocation and the resulting reductions in the economy's allocative efficiency and aggregate TFP depend on the distribution of wedges across the economy's production network. One consequence of focusing on a Cobb-Douglas economy with exogenous wedges, however, is that propagation patterns in the distorted and undistorted economies coincide. More precisely, if μ_k denotes the markup charged by industry k to all its customers, then

$$\frac{d}{d\mu_k} \left(\frac{d \log(y_i)}{d \log(z_j)} \right) = 0$$

for all i, j , and k ; i.e., the impact of a shock to industry j on i 's log output is invariant to the level of markups in the economy.

The interaction between distortions and productivities is investigated by Caliendo et al. (2018b), who consider a model of the world economy with CES technologies, and Baqaee & Farhi (2018b,c), who provide first-order approximations of the impact of productivity shocks in a fairly general class of economies while maintaining the wedges as exogenously given model primitives. Baqaee & Farhi (2018c) show that productivity shocks' first-order impact can be decomposed into two separate terms: (a) a term that accounts for the shocks' pure technology effect and (b) an additional term that accounts for changes in the economy's allocative efficiency. In addition, they show that, away from the Cobb-Douglas benchmark, the presence of distortions can change the productivity shocks' propagation patterns due to changes in the economy's allocative efficiency.

In a related study, Liu (2018) characterizes how market imperfections in the form of deadweight losses in input usage can result in misallocation of productive resources across sectors, thereby creating room for welfare-improving interventions in the form of industrial policies observed in some

¹²A related set of papers, such as those of Luo (2016), Altinoglu (2018), and Reischer (2018), studies how trade credit relationships shape the propagation of financial shocks over production networks.

developing countries. He demonstrates that, in the presence of input–output linkages, targeting the most distorted industries may be suboptimal, as market imperfections may accumulate in a nontrivial manner over the production network.¹³

While they are simple, reduced-form exogenous wedges or markups are not adequate for capturing how specific market imperfections shape propagation dynamics. Such an analysis requires a microfounded model of the interaction between shocks and wedges. Grassi (2017) takes a first step in this direction by considering a model of production networks with oligopolistic market structures. In such an environment, firm-level productivity shocks affect not only prices, but also markups via changes in the firms' competitiveness vis-à-vis other firms in the same industry. This means that the responsiveness of prices to shocks is no longer invariant to the realization of shocks, thus impacting the extent of downstream propagation. Furthermore, changes in market concentration impact industries' demand for intermediate inputs, thus inducing an upstream propagation channel that would be absent in a model with exogenous markups. In the same spirit, Baqaee (2018) endogenizes the mass of firms active in each industry in the context of an economy with imperfect competition and external economies of scale due to firm entry and exit. He shows that exits in an industry can change the profitability of firms in other industries and thus trigger endogenous adjustments in the mass of active firms. This creates an amplification channel in the form of upstream and downstream cascades of exits.

2.5. Endogenous Production Networks

Our discussion above is based on the assumption that, while input–output linkages can function as a shock propagation mechanism, the structure of the production network itself is invariant to the shocks. In reality, however, firms systematically respond to changes in economic conditions by altering their trading partners. For instance, they may source new inputs to take advantage of technological innovations or may enter into relationships with new customers in response to a customer's exit. Such endogenous changes in the production network can, in turn, significantly alter the economy's response to exogenous disturbances.

To accommodate the production network's response to shocks, a small but growing literature focuses on developing a joint theory of production and endogenous network formation. In developing such a theory, however, one faces a central challenge: The complexity inherent to direct and indirect network effects, coupled with the combinatorial nature of graphs, means that the relevant state space for firm-level decision making can become prohibitively large, even in fairly small economies consisting of a handful of firms.

One set of papers sidesteps this challenge by proposing statistical models of network formation. Atalay et al. (2011) develop a model in which links between firms are created through a variant of the preferential attachment model, while Carvalho & Voigtländer (2015) propose an industry-level network formation model based on the friendship model of Jackson & Rogers (2007), according to which existing input–output linkages are used to search for new inputs for production. Using industry-level data, and consistent with the model's central mechanism, they find that producers are more likely to adopt inputs that are already in use by their current (direct or indirect) upstream suppliers.

While statistical models like the ones mentioned above are able to match some of the key attributes of real-world production networks, by their nature, they abstract from firms' link-formation incentives. These incentives are explicitly incorporated by Oberfield (2018) into a

¹³ In the same spirit, King et al. (2018) study optimal carbon tax policies in a multisector economy with input–output linkages and show that the most effective carbon tax policy targets industries based not only on their individual emission levels, but also on their position in the production network.

dynamic model of network formation in which producers optimally choose one input from a randomly evolving set of suppliers. He finds that such endogenous choice results in the emergence of star suppliers that sell their goods to many other firms for intermediate use. Oberfield overcomes the curse of dimensionality discussed above by (a) considering an economy consisting of a continuum of firms and (b) restricting attention to single-input production technologies. These assumptions simplify the analysis by guaranteeing that each individual firm's decision has no impact on aggregate variables, and that equilibrium production networks that exhibit cycles are of measure zero.¹⁴

Acemoglu & Azar (2018) consider an alternative model in which firms in each one of n industries decide which subset of the other $n - 1$ industries to use as input suppliers, with each input combination leading to a different constant returns to scale production technology. The key assumption in the model is that markets are contestable in the sense that a large number of firms have access to the same menu of technologies. This assumption ensures that, when choosing its input combination, each firm can take the production network and all prices as given, thus bypassing complex strategic considerations of how its choice may reverberate through the network. In such an environment, aside from its standard effect of reducing all downstream prices (relative to the wage w), a positive technology shock to an industry alters the incentives of firms in downstream industries to adopt a wider set of inputs.

A related propagation mechanism is explored by Taschereau-Dumouchel (2018), who develops a firm-level model of network formation in which firms exit if they cannot meet fixed costs of production. Such extensive margin adjustments create strong complementarities between firms' operating decisions: A negative shock that results in a firm's exit reduces the profitability of its suppliers and customers, thus creating the potential for a cascade of shutdowns that changes the shape of the production network.¹⁵ The resulting propagation mechanism implies that periods of low economic activity feature a less clustered production network, a prediction that is consistent with the data.¹⁶

3. THE NETWORK ORIGINS OF AGGREGATE FLUCTUATIONS

The discussion in Section 2 illustrates that the economy's production network can function as a mechanism for propagating shocks from one firm or industry to the rest of the economy. Can such a propagation mechanism translate idiosyncratic microeconomic shocks into significant fluctuations at the aggregate level? The answer to this question can shed light on whether macroeconomic fluctuations can have their origins in idiosyncratic shocks to individual firms or disaggregated industries.

Going as far back as Lucas (1977), the possibility that significant fluctuations in aggregate economic variables may originate from microeconomic shocks was downplayed by the literature. This dismissal was based on a diversification argument, which maintained that, in an economy

¹⁴Lim (2018) adopts a similar approach by considering an economy consisting of a continuum of firms, but in contrast to Oberfield (2018), he allows for multi-input firms that are subject to relationship-specific cost shocks. The trade-off of these costs against the benefits of maintaining supplier–customer relationships generates extensive margin adjustments in firm-to-firm trade (see also Boehm & Oberfield 2018).

¹⁵The firm-level nature of the model, alongside the binary decision faced by the firms, makes the model analytically and computationally intractable. However, Taschereau-Dumouchel (2018) illustrates that, under certain conditions, the social planner's problem can be solved efficiently.

¹⁶A related set of papers, such as those of Antràs & Chor (2013), Chaney (2014), Antràs et al. (2017), and Tintelnot et al. (2018), studies firms' sourcing decisions in the international trade context. Chaney (2016), Johnson (2018), and Bernard & Moxnes (2018) provide a general overview of network models in international trade.

consisting of n industries hit by independent shocks, the standard deviation of aggregate fluctuations would be roughly proportional to $1/\sqrt{n}$, a negligible effect at high levels of disaggregation (corresponding to large values of n). This argument, however, ignores the possibility that shocks may propagate from one firm or industry to another over input–output linkages: With such a propagation mechanism at work, sectoral outputs would be correlated and thus may not wash out upon aggregation, even when the shocks themselves are independent.

In this section, we follow Acemoglu et al. (2012) and use Theorems 1 and 2 to revisit Lucas's (1977) argument and characterize the conditions under which input–output linkages in the economy can indeed generate sizable aggregate fluctuations from purely idiosyncratic shocks. We do so in two steps. We first use Equation 7 to relate the economy's aggregate volatility to the distribution of sectoral Domar weights in the economy. We then rely on Equation 8 to provide a characterization of network-originated macro fluctuations in terms of the economy's production network structure.

3.1. Micro Shocks and Macro Fluctuations

To illustrate the key ideas in the most transparent manner, we impose the following regularity assumptions on the baseline model from Section 2.1. First, we assume that log productivity shocks $\epsilon_i = \log(z_i)$ are independent and identically distributed with mean zero and finite standard deviation σ , thus ensuring that the economy is only subject to industry-level idiosyncratic shocks. Second, we suppose that the share of labor is the same across all industries, i.e., $\alpha_i = \alpha$ for all i .

Equation 7 implies that the aggregate volatility that is due to idiosyncratic microeconomic shocks is given by

$$\sigma_{\text{agg}} = \text{stdev}(\log(\text{GDP})) = \sigma \|\lambda\|, \quad 14.$$

where $\|\lambda\| = (\sum_{i=1}^n \lambda_i^2)^{1/2}$ denotes the second (uncentered) moment of Domar weights. Since Domar weights sum up to $\sum_{i=1}^n \lambda_i = \sum_{i,j=1}^n \beta_j \ell_{ji} = 1/\alpha$, we can rewrite the above equation as

$$\sigma_{\text{agg}} = \frac{\sigma/\alpha}{\sqrt{n}} \sqrt{1 + n^2 \alpha^2 \text{var}(\lambda_1, \dots, \lambda_n)}. \quad 15.$$

This relationship has two immediate implications. First, it implies that, when all Domar weights are identical, σ_{agg} is proportional to $1/\sqrt{n}$, consistent with the diversification argument. Second, the fact that, in general, σ_{agg} depends on the variance of Domar weights indicates that the argument put forth by Lucas (1977) may break down if sectoral Domar weights exhibit significant heterogeneity. In particular, Equation 15 illustrates that, all else equal, more dispersion in Domar weights results in higher levels of aggregate volatility emerging from purely idiosyncratic shocks.

These observations are at the heart of what Gabaix (2011, 2016) refers to as the granularity hypothesis: In the presence of significant heterogeneity at the micro level, the incompressible grains of economic activity (comprised of firms or disaggregated industries) can matter for the behavior of macroeconomic aggregates. This is driven by the fact that such heterogeneity reduces the extent to which various shocks cancel each other out at the aggregate level. Importantly, Gabaix (2011) also shows that, when the distribution of Domar weights is sufficiently heavy tailed, aggregate volatility can be significantly larger than Lucas's (1977) $1/\sqrt{n}$ benchmark, even at high levels of disaggregation. For example, suppose that Domar weights have a Pareto distribution with exponent $\gamma \geq 1$, in the sense that the fraction of industries with Domar weights greater than any given $\underline{\lambda}$ is proportional to $\underline{\lambda}^{-\gamma}$, with a smaller γ corresponding to more heterogeneity in

Domar weights. One can show that, when $\gamma \in (1, 2)$, $\|\lambda\|$ in Equation 14 is proportional to $n^{1/\gamma-1}$ as $n \rightarrow \infty$.¹⁷ Thus, a sufficiently skewed distribution of Domar weights can result in significantly higher levels of aggregate volatility relative to the $1/\sqrt{n}$ benchmark.

3.2. Network-Originated Macroeconomic Fluctuations

In this section, we turn to our main question of interest, namely, whether the economy's production network can translate idiosyncratic shocks into sizable macroeconomic fluctuations.

We first note that, while Equation 15 readily establishes that the extent of micro-originated GDP fluctuations is tightly linked to the heterogeneity in Domar weights, these weights are endogenous objects that are determined in equilibrium. We thus use Equation 8 to obtain a characterization in terms of the economy's structural parameters. Furthermore, to isolate the role of input–output linkages, we normalize the preference shares such that $\beta_i = 1/n$ for all i . Such a normalization ensures that any heterogeneity in Domar weights only reflects differences in the roles of different industries in the economy's production network. In particular, $\lambda_i = v_i/n$, where $v_i = \sum_{j=1}^n \ell_{ji}$ is the i th column sum of the economy's Leontief inverse and measures the importance of industry i as a direct or indirect input supplier to all sectors in the economy.

We can use Equations 8 and 15 to relate the volatility of log output of this economy to its production network:

$$\sigma_{\text{agg}} = \frac{\sigma}{\sqrt{n}} \sqrt{\alpha^{-2} + \text{var}(v_1, \dots, v_n)}. \quad 16.$$

Equation 16 recovers the key insight of Acemoglu et al. (2012): Sufficient heterogeneity in various industries' roles as input suppliers can lead to significantly higher levels of aggregate volatility compared to the $1/\sqrt{n}$ rate predicted by the diversification argument. For example, if the v_i s have a Pareto distribution with exponent $\gamma \in (1, 2)$, then σ_{agg} will be proportional to $n^{1/\gamma-1}$ as $n \rightarrow \infty$. The intuition underlying this result is tightly linked to the nature of the propagation mechanism as captured by Theorems 1 and 2. Microeconomic shocks wash out at the aggregate level if they impact the aggregate output roughly symmetrically. However, when industries are highly asymmetric in their roles as input suppliers, shocks to industries that are more important suppliers propagate more widely and thus do not wash out with the rest of the shocks upon aggregation.

To further clarify how input–output linkages may shape aggregate volatility, it is instructive to consider the graph-theoretic interpretations of vector $v = (v_1, \dots, v_n)$ and Equation 16. By its definition, the Leontief inverse satisfies $\mathbf{L} = \mathbf{I} + \mathbf{LA}$. Consequently, $v_i = \sum_{j=1}^n \ell_{ji}$ can be expressed in a recursive form as

$$v_i = 1 + \sum_{j=1}^n a_{ji} v_j.$$

This representation indicates that v_i coincides with the so-called Bonacich centrality of vertex i in the graph that represents the economy's production network (Bonacich 1987): An industry i is more central in the production network if it is a more important input supplier to other central industries. Thus, according to this interpretation, Equation 16 establishes that microeconomic shocks can generate sizable aggregate fluctuations when the economy's production network consists of industries with widely disparate centralities. **Figure 2a** provides an example of one such

¹⁷If $\gamma > 2$, then $\|\lambda\|$ scales as $1/\sqrt{n}$ as $n \rightarrow \infty$, whereas in the knife-edge case of $\gamma = 2$, it scales as $\sqrt{\log(n)/n}$ (for detailed derivations, see Gabaix 2011, proof of proposition 2; Acemoglu et al. 2017, proof of corollary 1).

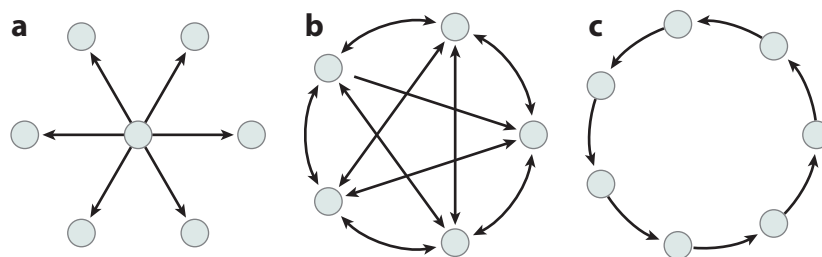


Figure 2

Production networks corresponding to three economies with nontrivial input–output linkages. Each vertex corresponds to an industry, with a directed edge present from one vertex to another if the former is an input supplier to the latter.

economy (albeit an extreme one), in which a single industry serves as the sole input supplier to all other industries. As such, microeconomic shocks to this central industry propagate widely throughout the economy and thus do not wash out with the rest of the shocks. In fact, it is easy to verify that, among all economies with the same labor share α , the star network in **Figure 2a** exhibits the maximal dispersion in industrial centralities, $\text{var}(v_1, \dots, v_n)$. The economies depicted in **Figure 2b,c** are at the other end of the spectrum, with all industries taking symmetric roles as input suppliers. Note that, despite the fact that micro shocks do propagate over these networks, the fact that they propagate symmetrically means that they will cancel each other out, leading to minimal aggregate effects. Indeed, Equation 16 implies that the volatility of aggregate fluctuations driven by idiosyncratic shocks in these economies is proportional to $1/\sqrt{n}$.¹⁸

3.3. Comovements

Our discussion above shows that sufficient heterogeneity in Domar weights can translate microeconomic shocks to highly disaggregated industries into macroeconomic fluctuations. Importantly, this is the case regardless of whether such heterogeneity is driven by asymmetry in the economy's production network (as in Section 3.2) or is due to other reasons [for example, heterogeneity in preference shares $(\beta_1, \dots, \beta_n)$]. By now, however, it should be clear that, even when the source of heterogeneity in Domar weights may not matter for aggregate fluctuations, economies that exhibit higher levels of network heterogeneity exhibit higher levels of comovement: Propagation of shocks over the economy's production network increases the likelihood that more industries move in tandem over the business cycle.

To formalize this statement, we consider the benchmark model from Section 2.1 under the restriction that the economy's input–output matrix is a symmetric circulant matrix with diagonals that are greater than $1/n$.¹⁹ While it is somewhat restrictive, the focus on this subclass of economies provides us with enough structure to present the key ideas in the most transparent manner. We maintain the assumption that microeconomic shocks $(\epsilon_1, \dots, \epsilon_n)$ are independent and identically distributed with mean zero and standard deviation σ .

¹⁸Whether input–output linkages can turn microeconomic shocks into sizable aggregate fluctuations is debated by Horvath (1998, 2000) and Dupor (1999), with the former arguing in favor, while the latter provided analytical results for a broad class of economies with fairly dense production networks—such as **Figure 2b**—in which micro shocks wash out at a fairly rapid rate upon aggregation. Our discussion above and Equation 16 illustrate that Dupor's results were driven by his focus on economies in which all industries have identical centralities (despite the presence of nontrivial input–output linkages).

¹⁹A matrix is said to be circulant if each row is a single-element rotation of the previous row.

Given two economies in this class with input–output matrices \mathbf{A} and $\tilde{\mathbf{A}}$ and identical Domar weights ($\lambda_i = \tilde{\lambda}_i$ for all i) and labor shares ($\alpha_i = \tilde{\alpha}_i = \alpha$ for all i), we say that the latter economy is more interconnected than the former if $\tilde{a}_{ij} = \gamma a_{ij} + (1 - \gamma)(1 - \alpha)/n$ for all pairs of industries i and j and some $\gamma \in [0, 1]$. Under this definition, the intensity of input–output linkages between any two industries in the more interconnected economy is more evenly distributed, with a small value of γ corresponding to an economy that is more similar to the complete network in **Figure 2b**.²⁰ We have the following novel result, the proof of which is provided in the **Supplemental Appendix**.

Theorem 4. Consider a pair of economies with identical Domar weights, and suppose that the latter is more interconnected than the former. Then,

1. the average pairwise correlation of (log) outputs is higher in the more interconnected economy, and
2. the industries in the less interconnected economy are more volatile.

Statement 1 of Theorem 4 thus formalizes our claim regarding the importance of input–output linkages in creating comovement across different industries: Given two economies with identical Domar weights, the one with higher levels of interconnectivity leads to higher average pairwise correlations, despite the fact that the two economies are indistinguishable at the aggregate level. Statement 2 then establishes that this increase in comovement is coupled with a reduction in sectoral volatilities. This is a consequence of the fact that, in a more interconnected economy with a more even distribution of input–output linkages, each industry is more diversified with respect to the upstream risk emanating from its suppliers, its suppliers’ suppliers, and so on.

Taken together, the two statements in Theorem 4 illustrate a key distinction between the nature of economic fluctuations in (a) an economy with high levels of network heterogeneity and (b) an economy with an identical Domar weight distribution but with low levels of network heterogeneity. Aggregate fluctuations in the latter economy are, for the most part, driven by fluctuations in sectors with high Domar weights. In contrast, sizable aggregate fluctuations that arise from the interplay of microeconomic shocks and the production network exhibit significant comovement across a wide range of sectors within the economy.

3.4. Macroeconomic Tail Risks

Our focus in the sections above is on how input–output linkages can shape (a) the economy’s aggregate volatility and (b) the comovement among various industries, as measured by their variance–covariance matrix. The economy’s production network, however, has implications for the distribution of sectoral and aggregate outputs well beyond their second moments.

These implications are the focus of Acemoglu et al. (2017), who show that the economy’s production network may fundamentally reshape the distribution of output by increasing the likelihood of large economic downturns from infinitesimal to substantial. Such an analysis requires a notion for measuring macroeconomic tail risks. A natural candidate is to measure macro tail risks in terms of systematic departures in the frequency of large contraction in aggregate output from what would prevail under a normal distribution with an identical variance. The central result is that an economy with a nontrivial production network that is subject to thin-tailed shocks may

²⁰While related, this notion of interconnectivity is distinct from the one defined by Acemoglu et al. (2017). The transformation $\tilde{a}_{ij} = \gamma a_{ij} + (1 - \gamma)(1 - \alpha)/n$ coincides with the concept of γ -convex combination of two networks in the work of Acemoglu et al. (2015).

exhibit deep recessions as frequently as economies that are subject to heavy-tailed shocks. Importantly, whereas the second moment of the distribution of Domar weights (i.e., $\|\lambda\|$ in Equation 14) is a sufficient statistic for the extent of micro-originated volatility, the extent of macroeconomic tail risks is determined by a statistic that also depends on the largest Domar weight in the economy. This disparity implies that the role of production networks in generating macroeconomic tail risks is distinct from their role in generating high levels of aggregate volatility. Thus, macroeconomic tail risks may vary significantly even across economies that exhibit otherwise identical behavior for moderate deviations.

4. EMPIRICAL AND QUANTITATIVE STUDIES

In the sections above, we use a simple model and a few of its variants to present the theoretical foundations for the role of production networks in macroeconomics. In this section, we provide a brief guide to the literature that explores these themes empirically and quantitatively.

4.1. Properties of Production Networks

We start with an overview of some of the well-documented stylized facts concerning firm- and industry-level production networks.

Perhaps the most widely used industry-level data are the Input–Output Accounts Data compiled by the Bureau of Economic Analysis (BEA). This database contains the most disaggregated sectoral data available worldwide, providing a detailed breakdown of the US economy into hundreds of industries. As documented by Carvalho (2010, 2014) and Acemoglu et al. (2012), among others, the BEA data indicate that the US industry-level production network exhibits a few notable properties. First, the industry-level network is highly sparsely connected, in the sense that narrowly defined specialized industries supply inputs to only approximately 11 other industries on average. Second, it is dominated by a small number of hubs: general-purpose industries that supply a wide range of industries in the economy. This is reflected in a highly skewed distribution of (weighted) outdegrees, which is well approximated by a Pareto distribution. Third, the production network exhibits what has come to be known as the small-world property, where, although most industry pairs are not directly linked by an input-supply relation, they are indirectly linked by hub-like sectors, resulting in a network with short average pairwise path length and a small diameter. Finally, the production network exhibits a highly skewed distribution of sectoral (Bonacich) centralities, also well approximated by a Pareto distribution with diverging second moments. As reviewed in Section 3.2, this property indicates that, in the data, heterogeneity in centrality is extensive enough to lead to a breakdown of the diversification argument, implying the possibility that micro shocks generate sizable aggregate fluctuations. As we discuss below, this possibility is indeed confirmed by a host of quantitative studies.

Industry-level input–output data are also available for many other countries, albeit at considerably coarser levels. The STAN database (containing benchmarked input–output data for 47 industries across 37 OECD countries) and the Global Trade Analysis database (with better coverage of low-income countries at a slightly higher level of aggregation) allow for cross-country comparative studies of production networks. Using these data, McNerney et al. (2013), Blöchl et al. (2011), and Fadinger et al. (2018) document that, consistent with the patterns in the United States, the distributions of sectoral outdegrees and centralities are highly heterogeneous in a wide range of countries. In addition, Blöchl et al. (2011) document that different groups of countries cluster around different central industries, while Fadinger et al. (2018) find that central industries in richer countries are relatively less productive.

The recent availability of large-scale firm-level transactions data has made analyses at a more granular level possible. One of the most extensive of such data sets originates from a large (private) credit reporting agency in Japan, Tokyo Shoko Research (TSR), which, in the course of issuing credit scores for firms, obtains the identity of firms' customers and suppliers. This yields information on the buyer-supplier relationships of close to 1 million firms, virtually covering the universe of Japanese firms with more than five employees. Another important source of firm-level transactions data is value-added tax (VAT) records from countries where such tax is levied and where tax authorities require the reporting of transactions between any two VAT-liable entities. The best studied of such data sets is based on VAT records from Belgium, containing the universe of all domestic supplier-customer relationships at the firm level. Relative to the Japanese data mentioned above, these data are richer, as they also contain the transaction amounts associated with each of the firm-to-firm links. Carvalho et al. (2016) and Bernard et al. (2019b) report some stylized facts emerging from the Japanese data, while Bernard et al. (2019a) do the same for Belgium. While they are based on two different countries, these studies suggest a number of salient characteristics of firm-level production networks. First, as in the case of industries, firm-level networks exhibit extensive heterogeneity in the role of firms as input suppliers, with outdegree distributions that are close to Pareto. Second, in contrast to industry-level networks, the indegree distributions are also very skewed, indicating the presence of firms that rely on a large number of suppliers. Third, larger firms in terms of sales or employees also tend to have larger numbers of buyers and suppliers. Finally, geographical distance is an important determinant in firm-to-firm link formation, with most linkages occurring between firms that are geographically close.

Unfortunately, the structure of the US firm-level production network has received less attention, as data are more scant. The most widely used US data set on buyer-supplier relations comes from the Compustat database, which is based on the financial accounting regulations that require publicly listed firms to disclose the identity of any customer representing more than 10% of their reported sales. Clearly, this induces a double selection bias: The data only contain linkages for publicly traded firms and typically correspond to small firms supplying to relatively larger customers. Nonetheless, the data can still provide valuable information about the nature of production at a granular level. For instance, Atalay et al. (2011) are able to document that, as in Japan and Belgium, the indegree distribution of the network of publicly listed firms in the United States is also highly skewed.²¹

4.2. Propagation Patterns

In this section, we review some of the empirical evidence on the propagation of shocks at the industry and firm levels.

4.2.1. Industry-level evidence. Acemoglu et al. (2016a) test the propagation mechanism implied by the baseline model in Section 2.1 at the industry level. Their starting point is the expression in Equation 6 for the equilibrium output of each industry as a function of the economy's production network and microeconomic productivity shocks. Taking first differences, this

²¹In addition to the Compustat database, Capital IQ, FactSet Revere, and Bloomberg also provide firm-level supply chain data for the United States (and several other countries) with an emphasis on publicly traded firms. Finally, as Atalay et al. (2014) show, another promising avenue for US data is to combine the micro survey data from the Commodity Flow Survey with the US manufacturing census to obtain information on plant-to-plant shipments of goods.

expression implies that

$$\Delta \log(y_i) = \Delta \epsilon_i + \sum_{j=1}^n (\ell_{ij} - \mathbb{I}_{(j=i)}) \Delta \epsilon_j,$$

where \mathbb{I} denotes the indicator function. The above expression decomposes the output growth of industry i into an own effect (the result of i 's own productivity shock, $\Delta \epsilon_i$) and a network effect due to the propagation of shocks from other industries. Acemoglu et al. operationalize this decomposition by constructing the Leontief inverse using the input–output tables compiled by the BEA and sourcing detailed sectoral output data from the NBER-CES manufacturing industry database. As a proxy for productivity shocks, they use lagged realizations of sector-level TFP growth so as to minimize concerns regarding contemporaneous joint determination of output and TFP. Combining input–output data with this candidate measure for shocks yields the main regressor of interest, defined as $\text{Downstream}_{i,t-1} = \sum_{j=1}^n (\ell_{ij} - \mathbb{I}_{(j=i)}) \Delta \text{TFP}_{j,t-1}$. This is a weighted average of shocks hitting i 's direct and indirect suppliers, using the entries of the Leontief inverse as weights, as instructed by the model. The labeling of this regressor reflects our discussion in Section 2.1 that, in a Cobb–Douglas economy, productivity shocks should only propagate downstream. This in turn implies that the corresponding upstream measure, $\text{Upstream}_{i,t-1} = \sum_{j=1}^n (\ell_{ji} - \mathbb{I}_{(j=i)}) \Delta \text{TFP}_{j,t-1}$, should have no effect on i 's output dynamics. The following regression can thus be used to test the propagation patterns implied by the baseline model:

$$\Delta \log(y_{it}) = \delta_t + \psi \Delta \log(y_{it-1}) + \beta_{\text{own}} \text{Own}_{it-1} + \beta_d \text{Downstream}_{it-1} + \beta_u \text{Upstream}_{it-1} + \varepsilon_{it},$$

where $\text{Own}_{i,t-1} = \Delta \text{TFP}_{i,t-1}$ captures industry i 's own direct productivity shock. This specification additionally allows for the presence of lagged dependent variables and year fixed effects to deal with possibly correlated error structures, either across time or in the cross-section.

Consistent with the theory, Acemoglu et al. (2016a) find that the downstream network effect of productivity shocks is economically and statistically significant: A one-standard-deviation increase in TFP growth is associated with a downstream effect of approximately 6% on output growth. By comparison, the upstream effect of productivity shocks is much smaller economically and its statistical significance is not robust to alternative output measures. These findings are in broad accordance with the predictions of Theorem 1 for the baseline Cobb–Douglas economy.

This simple empirical framework is flexible enough to also test the propagation patterns of demand shocks. Recall from Theorem 3 that, in the baseline Cobb–Douglas economy, demand-side shocks to a given industry should only propagate upstream to its direct and indirect suppliers. Acemoglu et al. (2016a) use changes in federal government spending to construct one such shock: They interact an industry's (initial) share of sales to the federal government with aggregate growth of federal spending. As in the previous exercise, the regressors of interest are obtained by lagging the shocks and constructing weighted averages of the shocks across direct and indirect suppliers (for upstream effects) and customers (for downstream effects), with weights given by the corresponding elements of the Leontief inverse. Consistent with the models' predictions, these results indicate significant upstream—rather than downstream—network effects following a demand shock.

We conclude by noting that, while the above exercise is indicative of propagation patterns that are broadly consistent with the predictions of Theorems 1 and 3, one should be careful in interpreting these estimates as causal. Even though the use of lagged predetermined shocks seeks to minimize endogeneity concerns, these shocks—and in particular TFP growth—may

be endogenous to decisions in the recent past that affect current realizations of the left- and right-hand side variables in the regression equations.

4.2.2. Firm-level evidence. A different strand of literature uses more granular data across a host of different countries to document the propagation of shocks at the firm level. These studies serve two important purposes. First, while industry-level evidence for the propagation of shocks indicates that production networks can have empirically relevant implications, ultimately, any actual propagation happens at the level of firms. Therefore, firm-level studies can provide more direct evidence for the nature of the underlying propagation mechanisms. Second, the possibility of identifying arguably exogenous shocks at the firm level (such as localized natural disasters), coupled with the more extensive variation in exposure to such shocks, means that one can overcome the endogeneity concerns that may arise at more aggregated levels.

Barrot & Sauvagnat (2016) investigate the propagation of firm-specific shocks by combining data on the timing and location of major natural disasters (in the form of blizzards, earthquakes, floods, and hurricanes) in the United States with information on the physical headquarters locations and supplier–customer linkages of publicly listed firms from Compustat. Given the limitations of the observable production network in the Compustat database, Barrot & Sauvagnat (2016) focus on local propagation patterns from a firm to its immediate suppliers and customers by regressing changes in quarterly sales of firms on a dummy variable capturing whether the firm's direct suppliers were located in a county hit by a natural disaster in a recent quarter. They document that exposures to the natural disaster results in a 2–3 percentage point drop in sales growth of the disrupted firm's direct customers. Importantly, this drop is particularly pronounced when the disrupted supplier is producing hard-to-substitute relation-specific inputs, in which case the shock further propagates to other (nonaffected) suppliers of the customer firm. This evidence suggests that, while the Cobb–Douglas model may serve as a good approximation at the industry level, it may break down at the more micro level, where easily substitutable inputs coexist with relation-specific inputs that are more difficult to substitute (at least in the short run).

A similar pattern is documented by Boehm et al. (2019), who use US Census Bureau microdata to study firm-level cross-country transmission of supply chain disruptions caused by the Great East Japan Earthquake of 2011. Combining reduced-form evidence with structural estimates of production elasticities, they find that the US affiliates of Japanese multinationals experienced a roughly one-for-one decline in output in response to declines in imports. This finding indicates that the short-run elasticity of substitution between imported and domestic inputs is close to zero.

While the above studies provide credible evidence for the propagation of shocks from a firm to its direct suppliers and customers, a shock's impact on the aggregate economy also depends on the extent to which it eventually propagates to more distant, only indirectly connected, firms. Testing this hypothesis, however, requires large-scale and detailed information on firm-to-firm linkages across the economy. This is the approach taken by Carvalho et al. (2016), who use the TSR data to trace the disruption caused by the 2011 earthquake and tsunami throughout the Japanese production network. Consistent with the results surveyed above, they find a significant postearthquake impact on the sales growth rates of firms with direct suppliers in the disaster areas. In addition, they also find that the disruption (a) propagated further downstream to the disaster area firms' indirect customers and (b) resulted in significant upstream propagation to the direct and indirect suppliers of earthquake-hit firms. The evidence on indirect propagation effects, coupled with the small-world nature of the production network, suggests that localized disturbances like the earthquake can have nontrivial aggregate consequences: While the individual, firm-level impact of the

disruption may not be very large—particularly when considering indirectly exposed firms—its aggregate effect can be significantly higher whenever a large fraction of firms in the economy is only two or three input links away from disrupted firms.

Understanding whether and how shocks propagate in production networks is currently the focus of a fast-expanding literature that combines novel production network data with a host of different shocks, thus going beyond the early interest in productivity disturbances. Demir et al. (2018) study the propagation and amplification of financial shocks by liquidity-constrained firms. Combining extensive VAT firm-to-firm transaction data from Turkey with an unexpected policy change levying a tax on trade credit financing by Turkish importers (which effectively made it more costly to finance input purchases from abroad), they find that liquidity-constrained importers exposed to the shock transmitted it to their downstream customers. Carvalho & Draca (2018) use detailed military procurement data from the US government and Compustat data on supply chain linkages for publicly listed firms to document that an increase in demand expands innovation efforts not only by final demand producers, but also by their upstream suppliers through recursive market size effects. Noting that expansionary monetary policy shocks—by acting as final demand shocks—should propagate upstream through the production network, Ozdagli & Weber (2017) investigate the role of such network effects as a possible transmission mechanism of monetary policy shocks. Finally, Auer et al. (2019) show that linkages across country–sector pairs contribute systematically to (producer price) inflation comovement across countries.²²

4.3. Comovements and Aggregate Fluctuations

In this section, we briefly survey the literature aimed at quantifying the role of production networks in generating comovement and aggregate fluctuations.

A reduced-form approach to accounting for industrial comovement is to appeal to a small number of common factors driving (correlated) output dynamics in many industries. Such an approach entails estimating a so-called approximate factor model on a panel of sectoral output growth rates, $\Delta \log y_t$, in the form of $\Delta \log y_t = \Lambda F_t + u_t$, where F_t is a low-dimensional vector of latent factors, Λ is a matrix of factor loadings of appropriate size, and u_t is a vector of industry-specific disturbances, assumed to satisfy weak cross-sectional dependence. The typical finding is that, from a reduced-form standpoint, a small number of common factors account for most comovement in data. However, recall from our discussion in Section 3.3 and Theorem 4 that production networks can induce significant comovement from purely idiosyncratic industry-specific shocks. Thus, what could appear to the econometrician as common shocks may instead be the result of endogenous comovement generated by the equilibrium interactions between various industries in a production network. Properly accounting for such a possibility calls for a structural approach that takes the input–output linkages explicitly into account.

Foerster et al. (2011) adopt one such structural approach to decompose the dynamics of disaggregated US industrial production indices into components arising from aggregate and industry-specific shocks. They use a dynamic variant of the baseline Cobb–Douglas economy in Section 2.1, featuring capital accumulation, capital goods’ linkages across industries, and more general preferences. By inverting the model-implied mapping from disturbances to observables to recover the underlying structural shocks, they find that idiosyncratic productivity shocks alone account for 50% of aggregate industrial production fluctuations between 1984 and 2007. Thus, while statistical models may perceive industrial comovement as being led almost solely by common macro

²²A smaller literature in finance investigates the asset pricing implications of production networks. Some recent examples include the work of Herskovic et al. (2017), Gofman et al. (2018), and Herskovic (2018).

shocks, a nontrivial fraction of aggregate fluctuations can instead be traced to idiosyncratic shocks propagating across the production network.²³

While important, these conclusions rely on the assumption that the structural model coincides with the true data-generating process and, as a result, on the particular propagation mechanics imparted by Cobb-Douglas technologies. Yet, as we discuss above, empirical micro studies of propagation patterns suggest important departures from the Cobb-Douglas benchmark, thus raising questions about the robustness of the quantitative inferences above. Atalay (2017) tackles this problem by showing how to extend the Foerster et al. (2011) methodology to an economy with CES technologies and preferences. To calibrate his model, Atalay (2017) uses annual input–output tables constructed by the BEA to estimate elasticities of substitution in industries’ production functions, obtaining a value of at most 0.2 for the elasticity of substitution between intermediate inputs. The strong complementarity among inputs suggested by these estimates implies stronger propagation and thus more pronounced aggregate effects originating from microeconomic shocks compared to the Cobb-Douglas benchmark. Indeed, under his benchmark parameter estimates, Atalay (2017) concludes that 83% of the variation in aggregate output growth is attributable to idiosyncratic industry-level shocks.

The estimates by Foerster et al. (2011) and Atalay (2017) are also in broad accordance with the earlier attempts of Horvath (2000) and Carvalho (2010) to quantify the macroeconomic importance of idiosyncratic shocks by directly calibrating large-scale multisector models under the assumption of uncorrelated disturbances. Both studies concluded that the interplay of idiosyncratic shocks with input–output linkages can account for approximately two-thirds of aggregate fluctuations.²⁴ Exploiting a different variance decomposition methodology, di Giovanni et al. (2014) find similarly sized effects for the contribution of linkages to aggregate volatility. Taken together, this body of work suggests that the economy’s production network is a major driver of comovements and GDP fluctuations.

Similarly to the empirical literature on propagation patterns surveyed above, there is an active interest in expanding the range of quantitative insights derived from taking production networks explicitly into account. Baqaee & Farhi (2018a) quantify the effects of CES-induced nonlinearities in production networks and show that such nonlinearities (*a*) amplify the effects of negative sectoral shocks while mitigating positive shocks; (*b*) generate significant negative skewness and excess kurtosis in aggregate output dynamics even when the underlying structural shocks are symmetric and thin tailed; and (*c*) can lead to significant welfare costs of business cycles, ranging from 0.2% to 1.3%, an order of magnitude larger than standard estimates in the literature. Bigio & La’O (2017) apply their model of production networks featuring Cobb-Douglas technologies and exogenous wedges to measure the impact of sectoral financial distortions during the Great Recession. They conclude that the production network amplified industry-level financial shocks from 1.7 to 2.4 times relative to an equivalent economy with no linkages. Grassi (2017) instead calibrates a model of interlinked oligopolistic market structures and finds that aggregate volatility arising from independent firm-level shocks accounts for 34% of what is observed in the data (for related firm-level calibration exercises, see Gabaix 2011, Carvalho & Grassi 2019). Relatedly, Magerman et al. (2017) and Kikkawa et al. (2019) exploit extensive Belgium VAT data on firm-to-firm trade to calibrate detailed models of firm-level production networks. They conclude, respectively, that firm-level

²³For comparison, Foerster et al. (2011) show that adopting the reduced-form approach of approximate factor models would lead one to conclude that two factors account for 87% of variability in aggregate industrial production between 1984 and 2007.

²⁴The reader is also referred to the related contributions of Shea (2002) and Conley & Dupor (2003) for earlier studies documenting aspects of sectoral comovements.

idiosyncratic shocks account for 57% of aggregate volatility and that firm-to-firm production networks entail a substantial amount of double marginalization, increasing by approximately 50% the welfare gains of reducing firm markups relative to a simpler roundabout economy featuring no network. In the same vein, Baqaee & Farhi (2018c) find that eliminating markup distortions entirely in an environment with production networks and CES production functions would raise aggregate TFP by 20%. Caliendo et al. (2018a) consider the interplay between production networks and the spatial structure of production. Their findings indicate that elasticities of TFP and GDP to regional and sectoral productivity changes vary significantly depending on the sectors and regions affected. Finally, Tintelnot et al. (2018) consider a quantitative model where domestic production networks coexist with international trade and where domestic firm-to-firm linkages can be endogenously rewired in response to international trade shocks. They find that allowing for the endogenous formation of the network in the model attenuates the costs of large negative trade shocks while amplifying the gains from trade following large positive ones.²⁵

5. CONCLUDING REMARKS

In this article, we provide a brief overview of the growing theoretical and empirical literature on the role of production networks in shaping economic outcomes. We rely on a simple benchmark model and several of its variants to illustrate how production networks can (a) function as a mechanism for the propagation of shocks throughout the economy and (b) translate microeconomic shocks into sizable fluctuations in macroeconomic aggregates. We also survey the literature that tests these mechanisms empirically and quantifies their implications. We conclude by discussing several open questions and promising avenues for future research.

While supplier–customer relationships that give rise to a production network are formed at the level of firms, most of the literature focuses on models that are better approximations to the nature of these interactions at the industry level. In particular, aside from a few exceptions discussed in the previous sections, the literature abstracts from important issues such as firm-specific relationships, market power, endogenous formation of supplier–customer linkages, and the possibility of firm failures. This is despite the fact that firm-level forces can have nontrivial implications for the micro and macro dynamics of production networks. Developing models that take such firm-level forces seriously can help capture the theoretical and empirical richness that is currently missing from the literature.

Relatedly, while the literature has mostly focused on how production networks can alter our understanding of the nature of business cycles, the implications for long-term growth have been left largely unexplored. A few studies, such as those of Ciccone (2002), Jones (2011), and Acemoglu & Azar (2018), have argued for the importance of input linkages for industrialization and long-run growth (see also Levine 2012, Bartelme & Gorodnichenko 2015). The development of richer firm-level models of production networks, coupled with the availability of ever more detailed data, can provide fruitful synergies with the resurgent literature on endogenous growth, which incorporates extensive heterogeneity at the micro level.

Finally, another promising avenue for future research is to investigate the implications of input–output linkages in environments featuring nominal rigidities. The contributions of Christiano (2016) and Pasten et al. (2018) already suggest that accounting for the network structure of production may result in quantitatively larger welfare costs of inflation, affect the slope of the Phillips curve, and alter the real effects of monetary policy. This, in turn, may have implications for the

²⁵ Relatedly, di Giovanni et al. (2018) conclude that the international trade linkages of French firms account for one-third of the comovement between France and the rest of the world.

design of optimal monetary policy. Exploring the theoretical and quantitative relevance of production networks for the conduct of monetary policy can be of first-order importance for policy makers.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We are grateful to Daron Acemoglu for various helpful conversations. V.M.C. acknowledges the financial support of the European Research Council grant 337054, the Philip Leverhulme Prize, and the Keynes Fellowship.

LITERATURE CITED

- Acemoglu D, Akcigit U, Kerr W. 2016a. Networks and the macroeconomy: an empirical exploration. *NBER Macroecon. Annu.* 30:276–335
- Acemoglu D, Azar PD. 2018. *Endogenous production networks*. NBER Work. Pap. 24116
- Acemoglu D, Carvalho VM, Ozdaglar A, Tahbaz-Salehi A. 2012. The network origins of aggregate fluctuations. *Econometrica* 80:1977–2016
- Acemoglu D, Ozdaglar A, Tahbaz-Salehi A. 2015. Systemic risk and stability in financial networks. *Am. Econ. Rev.* 105:564–608
- Acemoglu D, Ozdaglar A, Tahbaz-Salehi A. 2016b. Networks, shocks, and systemic risk. In *The Oxford Handbook on the Economics of Networks*, ed. Y Bramoullé, A Galeotti, B Rogers, pp. 569–607. Oxford, UK: Oxford Univ. Press
- Acemoglu D, Ozdaglar A, Tahbaz-Salehi A. 2017. Microeconomic origins of macroeconomic tail risks. *Am. Econ. Rev.* 107:54–108
- Altinoglu L. 2018. *The origins of aggregate fluctuations in a credit network economy*. Finance Econ. Discuss. Pap. 2018-031, Board Gov. Fed. Reserve Syst., Washington, DC
- Antràs P, Chor D. 2013. Organizing the global value chain. *Econometrica* 81:2127–204
- Antràs P, Fort TC, Tintelnot F. 2017. The margins of global sourcing: theory and evidence from U.S. firms. *Am. Econ. Rev.* 107:2514–64
- Atalay E. 2017. How important are sectoral shocks? *Am. Econ. J. Macroecon.* 9:254–80
- Atalay E, Hortaçsu A, Roberts J, Syverson C. 2011. Network structure of production. *PNAS* 108:5199–202
- Atalay E, Hortaçsu A, Syverson C. 2014. Vertical integration and input flows. *Am. Econ. Rev.* 104:1120–48
- Auer R, Levchenko AA, Saure P. 2019. International inflation spillovers through input linkages. *Rev. Econ. Stat.* In press
- Baqee DR. 2018. Cascading failures in production networks. *Econometrica* 86:1819–38
- Baqee DR, Farhi E. 2018a. *The macroeconomic impact of microeconomic shocks: beyond Hulten's theorem*. NBER Work. Pap. 23145
- Baqee DR, Farhi E. 2018b. *Macroeconomics with heterogeneous agents and input-output networks*. NBER Work. Pap. 24684
- Baqee DR, Farhi E. 2018c. *Productivity and misallocation in general equilibrium*. NBER Work. Pap. 24007
- Barrot JN, Sauvagnat J. 2016. Input specificity and the propagation of idiosyncratic shocks in production networks. *Q. J. Econ.* 131:1543–92
- Bartelme D, Gorodnichenko Y. 2015. *Linkages and economic development*. NBER Work. Pap. 21251
- Benhabib J, Nishimura K. 1979. The Hopf bifurcation and the existence and stability of closed orbits in multisector models of optimal economic growth. *J. Econ. Theory* 21:421–44

- Berman A, Plemmons RJ. 1979. *Nonnegative Matrices in the Mathematical Sciences*. New York: Academic Press
- Bernard AB, Dhyne E, Magerman G, Manova K, Moxnes A. 2019a. *The origins of firm heterogeneity: a production network approach*. NBER Work. Pap. 25441
- Bernard AB, Moxnes A. 2018. Networks and trade. *Annu. Rev. Econ.* 10:65–85
- Bernard AB, Moxnes A, Saito YU. 2019b. Production networks, geography, and firm performance. *J. Political Econ.* In press
- Bigio S, La'O J. 2017. *Distortions in production networks*. Work. Pap., Columbia Univ., New York
- Blöchl F, Theis FJ, Vega-Redondo F, Fisher EO. 2011. Vertex centralities in input-output networks reveal the structure of modern economies. *Phys. Rev. E* 83:046127
- Boehm CE, Flaaen A, Pandalai-Nayar N. 2019. Input linkages and the transmission of shocks: firm-level evidence from the 2011 Tohoku Earthquake. *Rev. Econ. Stat.* In press
- Boehm J, Oberfield E. 2018. *Misallocation in the market for inputs: enforcement and the organization of production*. NBER Work. Pap. 24937
- Bonacich P. 1987. Power and centrality: a family of measures. *Am. J. Sociol.* 92:1170–82
- Caliendo L, Parro F, Rossi-Hansberg E, Sarte PD. 2018a. The impact of regional and sectoral productivity changes on the U.S. economy. *Rev. Econ. Stud.* 85:2042–96
- Caliendo L, Parro F, Tsyvinski A. 2018b. *Distortions and the structure of the world economy*. NBER Work. Pap. 23332
- Carvalho VM. 2010. *Aggregate fluctuations and the network structure of intersectoral trade*. Work. Pap. 1206, Dep. Econ. Bus., Univ. Pompeu Fabra, Barcelona, Spain
- Carvalho VM. 2014. From micro to macro via production networks. *J. Econ. Perspect.* 28:23–48
- Carvalho VM, Draca M. 2018. *Cascading innovation*. Work. Pap., Univ. Cambridge, UK
- Carvalho VM, Gabaix X. 2013. The great diversification and its undoing. *Am. Econ. Rev.* 103:1697–727
- Carvalho VM, Grassi B. 2019. Large firm dynamics and the business cycle. *Am. Econ. Rev.* In press
- Carvalho VM, Nirei M, Saito Y, Tahbaz-Salehi A. 2016. *Supply chain disruptions: evidence from the Great East Japan Earthquake*. Work. Pap. 2017-01, Becker Friedman Inst. Econ., Univ. Chicago
- Carvalho VM, Voigtländer N. 2015. *Input diffusion and the evolution of production networks*. NBER Work. Pap. 20025
- Chaney T. 2014. The network structure of international trade. *Am. Econ. Rev.* 104:3600–34
- Chaney T. 2016. Networks in international trade. In *Oxford Handbook of the Economics of Networks*, ed. Y Bramoullé, A Galeotti, B Rogers, pp. 754–75. Oxford, UK: Oxford Univ. Press
- Christiano LJ. 2016. Comment on “Networks and the macroeconomy: an empirical exploration.” *NBER Macroecon. Annu.* 30:346–73
- Ciccone A. 2002. Input chains and industrialization. *Rev. Econ. Stud.* 63:565–87
- Conley TG, Dopor B. 2003. A spatial analysis of sectoral complementarity. *J. Political Econ.* 111:311–52
- Demir B, Javorcik B, Michalsk TK, Örs E. 2018. *Financial constraints and propagation of shocks in production networks*. Work. Pap., Univ. Oxford, UK
- di Giovanni J, Levchenko AA, Méjean I. 2014. Firms, destinations, and aggregate fluctuations. *Econometrica* 82:1303–40
- di Giovanni J, Levchenko AA, Méjean I. 2018. The micro origins of international business cycle comovement. *Am. Econ. Rev.* 108:82–108
- Dopor B. 1999. Aggregation and irrelevance in multi-sector models. *J. Monet. Econ.* 43:391–409
- Fadinger H, Ghiglino C, Teteryatnikova M. 2018. *Income differences and input-output structure*. Work. Pap. 1510, Dep. Econ., Univ. Vienna
- Foerster AT, Sarte PDG, Watson MW. 2011. Sectoral versus aggregate shocks: a structural factor analysis of industrial production. *J. Political Econ.* 119:1–38
- Gabaix X. 2011. The granular origins of aggregate fluctuations. *Econometrica* 79:733–72
- Gabaix X. 2016. Comment on “Networks and the macroeconomy: an empirical exploration.” *NBER Macroecon. Annu.* 30:336–45
- Gofman M, Segal G, Wu Y. 2018. *Production networks and stock returns: the role of vertical creative destruction?* Work. Pap., Univ. Rochester, NY

- Grassi B. 2017. *IO in I-O: size, industrial organization, and the input-output network make a firm structurally important*. Work. Pap., Bocconi Univ., Milan, Italy
- Herskovic B. 2018. Networks in production: asset pricing implications. *J. Finance* 73:1785–818
- Herskovic B, Kelly B, Lustig H, Van Nieuwerburgh S. 2017. *Firm volatility in granular networks*. Res. Pap. 12–56, Booth School Bus., Univ. Chicago
- Horvath M. 1998. Cyclicalities and sectoral linkages: aggregate fluctuations from sectoral shocks. *Rev. Econ. Dyn.* 1:781–808
- Horvath M. 2000. Sectoral shocks and aggregate fluctuations. *J. Monet. Econ.* 45:69–106
- Hulten CR. 1978. Growth accounting with intermediate inputs. *Rev. Econ. Stud.* 45:511–18
- Jackson MO, Rogers BW. 2007. Meeting strangers and friends of friends: How random are social networks? *Am. Econ. Rev.* 97:890–915
- Johnson RC. 2018. Measuring global value chains. *Annu. Rev. Econ.* 10:207–36
- Jones CI. 2011. Intermediate goods and weak links: a theory of economic development. *Am. Econ. J. Macroecon.* 3:1–28
- Jones CI. 2013. Misallocation, economic growth, and input-output economics. In *Proceedings of Econometric Society World Congress*, ed. D Acemoglu, M Arellano, E Dekel, pp. 419–55. Cambridge, UK: Cambridge Univ. Press
- Kikkawa AK, Magerman G, Dhyne E. 2019. *Imperfect competition in firm-to-firm trade*. Work. Pap. 363, Natl. Bank Belgium, Brussels
- King M, Tarbush B, Teytelboym A. 2018. *Targeted carbon tax reforms*. Work. Pap., Univ. Oxford, UK
- Leontief WW. 1941. *The Structure of American Economy, 1919–1929: An Empirical Application of Equilibrium Analysis*. Cambridge, MA: Harvard Univ. Press
- Levine DK. 2012. Production chains. *Rev. Econ. Dyn.* 15:271–82
- Lim K. 2018. *Endogenous production networks and the business cycle*. Work. Pap., Univ. Toronto, Can.
- Liu E. 2018. *Industrial policies in production networks*. Work. Pap., Princeton Univ., Princeton, NJ
- Long JB, Plosser CI. 1983. Real business cycles. *J. Political Econ.* 91:39–69
- Lucas RE. 1977. Understanding business cycles. *Carnegie-Rochester Conf. Ser. Public Policy* 5:7–29
- Luo S. 2016. *Propagation of financial shocks in an input-output economy with trade and financial linkages of firms*. Work. Pap., Columbia Univ., New York
- Magerman G, Bruyne KD, Dhyne E, Van Hove J. 2017. *Heterogeneous firms and the micro origins of aggregate fluctuations*. Work. Pap. 312, Natl. Bank Belgium, Brussels
- McNerney J, Fath BD, Silverberg G. 2013. Network structure of interindustry flows. *Phys. A* 392:6427–41
- Oberfield E. 2018. A theory of input-output architecture. *Econometrica* 86:559–89
- Ozdagli A, Weber M. 2017. *Monetary policy through production networks: evidence from the stock market*. Work. Pap., Fed. Reserve Bank Boston
- Pasten E, Schoenle R, Weber M. 2018. *The propagation of monetary policy shocks in a heterogeneous production economy*. NBER Work. Pap. 25303
- Plemmons RJ. 1977. M-matrix characterizations. I: nonsingular M-matrices. *Linear Algebra Appl.* 18:175–88
- Reischer M. 2018. *Finance-thy-neighbor: trade credit origins of aggregate fluctuations*. Work. Pap., Univ. Cambridge, UK
- Shea J. 2002. Complementarities and comovements. *J. Money Credit Banking* 34:412–33
- Stewart GW. 1998. *Matrix Algorithms*, Vol. 1: *Basic Decompositions*. Philadelphia, PA: Soc. Indust. Appl. Math.
- Summers L. 2013. *Crises yesterday and today*. Speech at 14th Jacques Polak Annual IMF Research Conference in Honor of Stanley Fischer, Washington, DC, Nov. 7–8. <http://larrysummers.com/imf-fourteenth-annual-research-conference-in-honor-of-stanley-fischer>
- Taschereau-Dumouchel M. 2018. *Cascades and fluctuations in an economy with an endogenous production network*. Work. Pap., Cornell Univ., Ithaca, NY
- Tintelnot F, Kikkawa K, Mogstad M, Dhyne E. 2018. *Trade and domestic production networks*. NBER Work. Pap. 25120
- von Neumann J. 1945. A model of general economic equilibrium. *Rev. Econ. Stud.* 13:1–9