

THE MACROECONOMIC IMPACT OF MICROECONOMIC SHOCKS: BEYOND HULTEN'S THEOREM

DAVID REZZA BAQAEI
Department of Economics, UCLA

EMMANUEL FARHI
Department of Economics, Harvard University

We provide a nonlinear characterization of the macroeconomic impact of microeconomic productivity shocks in terms of reduced-form nonparametric elasticities for efficient economies. We also show how microeconomic parameters are mapped to these reduced-form general equilibrium elasticities. In this sense, we extend the foundational theorem of [Hulten \(1978\)](#) beyond the first order to capture nonlinearities. Key features ignored by first-order approximations that play a crucial role are: structural microeconomic elasticities of substitution, network linkages, structural microeconomic returns to scale, and the extent of factor reallocation. In a business-cycle calibration with sectoral shocks, nonlinearities magnify negative shocks and attenuate positive shocks, resulting in an aggregate output distribution that is asymmetric (negative skewness), fat-tailed (excess kurtosis), and has a negative mean, even when shocks are symmetric and thin-tailed. Average output losses due to short-run sectoral shocks are an order of magnitude larger than the welfare cost of business cycles calculated by [Lucas \(1987\)](#). Nonlinearities can also cause shocks to critical sectors to have disproportionate macroeconomic effects, almost tripling the estimated impact of the 1970s oil shocks on world aggregate output. Finally, in a long-run growth context, nonlinearities, which underpin Baumol's cost disease via the increase over time in the sales shares of low-growth bottleneck sectors, account for a 20 percentage point reduction in aggregate TFP growth over the period 1948–2014 in the United States.

KEYWORDS: Nonlinearities, disaggregated macro models, production networks, Hulten's theorem, Domar aggregation, Baumol's cost disease, oil shocks, welfare costs of business cycles.

1. INTRODUCTION

THE FOUNDATIONAL THEOREM of [Hulten \(1978\)](#) states that for efficient economies and under minimal assumptions, the impact on aggregate TFP of a microeconomic TFP shock is equal to the shocked producer's sales as a share of GDP:

$$d \log \text{TFP} = \sum_i \lambda_i d \log A_i,$$

where $d \log A_i$ is a shock to producer i and λ_i is its sales share or Domar weight.

Hulten's theorem is a cornerstone of productivity and growth accounting: it shows how to construct aggregate TFP growth from microeconomic TFP growth, and provides structurally-interpretable decompositions of changes of national or sectoral aggregates

David Rezza Baqaee: baqaee@econ.ucla.edu

Emmanuel Farhi: efarhi@harvard.edu

We thank Daron Acemoglu, Daniel Albuquerque, Pol Antras, Costas Arkolakis, Enghin Atalay, Natalie Bau, Francesco Caselli, Vasco Carvalho, Xavier Gabaix, Ben Golub, Charles Hulten, Dale Jorgenson, Adrien Fabre, John Moore, Ricardo Reis, Alireza Tahbaz-Salehi, and Jean Tirole for useful comments. We are grateful to Maria Voronina and Chang He for outstanding research assistance.

into the changes of their disaggregated component industries or firms. It also provides the benchmark answers for counterfactual questions in structural models with disaggregated production.

The surprising generality of the result has led economists to de-emphasize the role of microeconomic and network production structures in macroeconomic models. After all, if sales summarize the macroeconomic impact of microeconomic shocks and we can directly observe sales, then we need not concern ourselves with the details of the underlying disaggregated system that gave rise to these sales. Since it seems to imply that the very object of its study is irrelevant for macroeconomics, Hulten's theorem has been something of a bugbear for the burgeoning literature on production networks.

Are these conclusions warranted? Even at a purely intuitive level, there are reasons to be skeptical. Take, for example, shocks to Walmart and to electricity production. Both Walmart and electricity production have a similar sales share of roughly 4% of U.S. GDP. It seems natural to expect that a large negative shock to electricity production would be much more damaging than a similar shock to Walmart. Indeed, this intuition will be validated by our formal results. Yet it goes against the logic of Hulten's theorem which implies that, because the two sectors have the same Domar weight, the two shocks should have the same impact on aggregate output.

In this paper, we challenge the view that the macroeconomic importance of a microeconomic sector is summarized by its sales share and, more broadly, the notion that the microeconomic details of the production structure are irrelevant for macroeconomics. The key is to recognize that Hulten's theorem only provides a first-order approximation. Nonlinearities can significantly degrade the quality of the first-order approximation for large enough shocks. To capture these nonlinearities, we provide a general second-order approximation by characterizing the derivatives of Domar weights with respect to shocks. The second-order terms are shaped by the microeconomic details of the disaggregated production structure: network linkages, microeconomic elasticities of substitution in production, microeconomic returns to scale, and the degree to which factors can be reallocated.

Our results are general in that they apply to any efficient general equilibrium economy. They suggest that Cobb–Douglas models, commonly used in the production-network, growth, and multi-sector macroeconomics literatures, are very special: the Domar weights, and more generally the whole input–output matrix, are constant and can be taken to be exogenous, the first-order approximation is exact, the model is log-linear, and as a result, the microeconomic details of the production structure are irrelevant.¹ These knife-edge properties disappear as soon as one deviates from Cobb–Douglas: the Domar weights, and more generally the whole input–output matrix, respond endogenously to shocks, and the resulting nonlinearities are shaped by the microeconomic details of the production structure.

We also show that nonlinearities in production matter quantitatively for a number of macroeconomic phenomena operating at different frequencies, ranging from the role of sectoral shocks in business cycles to the impact of oil shocks and the importance of Baumol's cost disease for long-run growth:

1. Using a calibrated structural multi-industry model with realistic complementarities in production, we find that nonlinearities amplify the impact of negative sectoral shocks

¹A mixture of analytical tractability, as well as balanced-growth considerations, have made Cobb–Douglas the canonical production function for networks (Long and Plosser (1983)), multi-sector RBC models (Gomme and Rupert (2007)), and growth theory (Aghion and Howitt (2008)). Recent work by Grossman, Helpman, Oberfield, Sampson (2016) shows how balanced growth can occur without Cobb–Douglas.

and mitigate the impact of positive sectoral shocks.^{2,3} Large negative shocks to crucial industries, like “oil and gas,” have a significantly larger negative effect on aggregate output than negative shocks to larger but less crucial industries such as “retail trade.” Nonlinearities also have a significant impact on the distribution of aggregate output: they lower its mean and generate negative skewness and excess kurtosis even though the underlying shocks are symmetric and thin-tailed. Nonlinearities in production generate significant welfare costs of sectoral fluctuations, ranging from 0.2% to 1.3% depending on the calibration. These are an order of magnitude larger than the welfare costs of business cycles arising from nonlinearities in utility (risk aversion) identified by Lucas (1987).

2. We derive and use a simple nonparametric formula, taking into account the observed change in the Domar weight for crude oil, to analyze the impact of the energy crisis of the 1970s up to the second order. We find that nonlinearities almost tripled the impact of the oil shocks from 0.23% to 0.61% of world aggregate output.

3. We show that the nonlinearities are also important for long-run growth in the presence of realistic complementarities across sectors. They cause the Domar weights of bottleneck sectors with relatively low productivity growth to grow over time and thereby reduce aggregate growth, an effect identified as Baumol’s cost disease (Baumol (1967)). We calculate that nonlinearities have reduced the growth of aggregate TFP by 20 percentage points over the period 1948–2014 in the United States.⁴

The outline of the paper is as follows. In Section 2, we derive a general formula describing the second-order impact on aggregate output of shocks in terms of nonparametric sufficient statistics: reduced-form general equilibrium elasticities of substitution and input–output multipliers.⁵ We explain the implications of this formula for the impact of correlated shocks and for the average performance of the economy. In Section 3, we use two special illustrative examples to provide some intuition for the roles of the general equilibrium elasticities of substitution and of the input–output multipliers and for their dependence on microeconomic primitives. In Section 4, we fully characterize second-order terms in terms of microeconomic primitives for general nested-CES economies

²The empirical literature on production networks, like Atalay (2017), Boehm, Flaaen, and Pandalai-Nayar (2017), and Barrot and Sauvagnat (2016), all find that structural elasticities of substitution in production are significantly below 1, and sometimes very close to zero, across intermediate inputs, and between intermediate inputs and labor at business-cycle frequencies. Furthermore, a voluminous literature on structural transformation, building on Baumol (1967), has found evidence in favor of non-unitary elasticities of substitution in consumption and production across sectors over the long run.

³While complementarities prevail at the sectoral level, substitutabilities dominate across firms within sectors. This implies that while nonlinearities tend to amplify negative sectoral-level shocks and to attenuate positive sectoral-level shocks, they tend to attenuate negative firm-level shocks and to amplify positive firm-level shocks. Nonlinearities therefore introduce an important qualitative difference between sectoral- and firm-level shocks which is absent from the linearized perspective.

⁴The literature on structural transformation emphasizes two key forces: non-unitary elasticities of substitution and non-homotheticities. Both forces cause sales shares to change in response to exogenous shocks. Since Hulten’s theorem implies that sales shares are equal to derivatives of the aggregate output function, anything that causes the derivative to change is a nonlinearity. By characterizing the second-order terms in a general way, our results encompass both non-unitary elasticities and non-homotheticities. In fact, non-homotheticities can *always* be turned into non-unitary elasticities of substitution by adding more fixed-factors to an economy.

⁵Studying the second-order terms is the first step in grappling with the nonlinearities inherent in multi-sector models with production networks. In this sense, our work illustrates the macroeconomic importance of *local* and strongly *nonlinear* interactions emphasized in reduced form by Scheinkman and Woodford (1994). Other related work on nonlinear propagation of shocks in economic networks includes Durlauf (1993), Jovanovic (1987), Ballester, Calvo-Armengol, and Zenou (2006) Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015), Elliott et al. (2014), and especially Acemoglu, Ozdaglar, and Tahbaz-Salehi (2016).

with arbitrary microeconomic elasticities of substitution and network linkages. In Section 5, we further generalize the results to arbitrary (potentially non-CES) production functions. In Section 6, we provide some illustrations of the quantitative implications of our results.

Related Literature

Gabaix (2011) used Hulten's theorem to argue that the existence of very large, or in his language *granular*, firms can be a possible source of aggregate volatility. If there exist very large firms, then shocks to those firms will not cancel out with shocks to much smaller firms, resulting in aggregate fluctuations. Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012), working with a Cobb–Douglas model in the spirit of Long and Plosser (1983), observed that in an economy with input–output linkages, the equilibrium sizes of firms depend on the shape of the input–output matrix. Central suppliers will be weighted more highly than peripheral firms, and therefore, shocks to those central players will not cancel out with shocks to small firms.⁶ Carvalho and Gabaix (2013) showed how Hulten's theorem can be operationalized to decompose the sectoral sources of aggregate volatility.⁷

Relatedly, Acemoglu, Ozdaglar, and Tahbaz-Salehi (2017) deployed Hulten's theorem to study other moments of the distribution of aggregate output. They argued that if the Domar weights are fat-tailed and if the underlying idiosyncratic shocks are fat-tailed, then aggregate output can exhibit non-normal behavior. Stated differently, they showed that aggregate output can inherit tail risk from idiosyncratic tail risk if the distribution of the Domar weights is fat-tailed. Our paper makes a related but distinct point. We find that, for the empirically relevant range of parameters, the response of aggregate output to shocks is significantly asymmetric. Therefore, the nonlinearity inherent in the production structure can turn even symmetric thin-tailed sectoral shocks into rare disasters endogenously. This means that the economy could plausibly experience aggregate tail risk without either fat-tailed shocks or fat-tailed Domar weights.

In a recent survey article, Gabaix (2016), invoking Hulten's theorem, wrote “networks are a particular case of granularity rather than an alternative to it.” This has meant that researchers studying the role of networks have either moved away from efficient models, or that they have retreated from studying aggregate output and turned their attention to the microeconomic implications of networks, namely, the covariance of fluctuations between different industries and firms.^{8,9} However, our paper shows that except in very

⁶A related version of this argument was also advanced by Horvath (1998), who explored this issue quantitatively with a more general model in Horvath (2000). Separately, Carvalho (2010) also explored how the law of large numbers may fail under certain conditions on the input–output matrix.

⁷Results related to Hulten's theorem are also used in international trade, for example, Burstein and Cravino (2015), to infer the global gains from international trade.

⁸Some recent papers have investigated aggregate volatility in production networks with inefficient equilibria (where Hulten's theorem does not hold). Some examples include Bigio and La'O (2016), Baqaee (2018), Altinoglu (2016), Grassi (2017), and Baqaee and Farhi (2018c). See also Jones (2011, 2013), Bartelme and Gorodnichenko (2015), and Liu (2017).

⁹Other papers investigate the importance of idiosyncratic shocks propagating through networks to generate cross-sectional covariances, but refrain from analyzing aggregate output. Some examples include Foerster, Sarte, and Watson (2011), Atalay (2017), Di Giovanni, Levchenko, and Méjean (2014), Stella (2015), and Baqaee and Farhi (2018b). Atalay (2017) is particularly relevant in this context, since he found that structural elasticities of substitution in production play a powerful role in generating covariance in sectoral output. Our paper complements this analysis by focusing instead on the way complementarities affect aggregate output.

special cases, models with the same sales distributions but different network structures only have the same aggregate-output implications up to a fragile first order of approximation. Their common sales distributions produce the same linearization, but their different network structures lead to different nonlinearities. Hence, in the context of aggregate fluctuations, networks are neither a particular case of granularity nor an alternative to it. It is simply that the sales distribution is a sufficient statistic for the network at the first order but not at higher orders.

2. GENERAL FRAMEWORK

In this section, we set up a nonparametric general equilibrium model to demonstrate both Hulten's theorem as well as our second-order approximation. Final demand is represented as the maximizer of a constant-returns aggregator of final demand for individual goods

$$Y = \max_{\{c_1, \dots, c_N\}} \mathcal{D}(c_1, \dots, c_N),$$

subject to the budget constraint

$$\sum_{i=1}^N p_i c_i = \sum_{f=1}^F w_f \bar{l}_f + \sum_{i=1}^N \pi_i,$$

where c_i is the representative household's consumption of good i , p_i is the price and π_i is the profit of producer i , w_f is the wage of factor f which is in fixed supply \bar{l}_f . The two sides of the budget constraint coincide with nominal GDP, using respectively the final expenditure and income approaches.

Each good i is produced by competitive firms using the production function

$$y_i = A_i F_i(l_{i1}, \dots, l_{iF}, x_{i1}, \dots, x_{iN}),$$

where A_i is Hicks-neutral technology, x_{ij} are intermediate inputs of good j used in the production of good i , and l_{if} is labor type f used by i . The profits earned by the producer of good i are

$$\pi_i = p_i y_i - \sum_{f=1}^F w_f l_{if} - \sum_{j=1}^N p_j x_{ij}.$$

The market-clearing conditions for goods $1 \leq i \leq N$ and factors $1 \leq f \leq F$ are

$$y_i = \sum_{j=1}^N x_{ji} + c_i \quad \text{and} \quad \bar{l}_f = \sum_{i=1}^N l_{if}.$$

Competitive equilibrium is defined in the usual way, where all agents take prices as given, and markets for every good and every type of labor clear.

We interpret Y as a cardinal measure of (real) aggregate output and note that it is the correct measure of the household's "standard of living" in this model. We implicitly rely on the existence of complete financial markets and homotheticity of preferences to ensure the existence of a representative consumer. Although the assumption of a representative

consumer is not strictly necessary for the results in this section, it is a standard assumption in this literature since it allows us to unambiguously define and measure changes in real aggregate output without contending with the issue of the appropriate price index.

We assume that the production function F_i of each good i has constant returns to scale, which implies that equilibrium profits are zero. This assumption is less restrictive than it may appear because decreasing returns to scale can be captured by adding fixed factors to which the corresponding profits accrue.¹⁰ A similar observation applies to the assumption that shocks are Hicks-neutral: we can represent a productivity shock augmenting a specific input by adding a new producer that produces this input and hitting this new producer with a Hicks-neutral shock.¹¹ Note also that although we refer to each producer as producing one good, our framework actually allows for joint production by multi-product producers: for example, to capture a producer i producing goods i and i' using intermediate inputs and factors, we represent good i' as an input entering negatively in the production and cost functions for good i .¹² Finally, note that goods could represent different varieties of goods from the same industry, goods from different industries, or even goods in different time periods, regions, or states of nature.¹³

Define $Y(A_1, \dots, A_N)$ to be the equilibrium aggregate output as a function of the exogenous technology levels. Throughout the paper, and without loss of generality, we derive results regarding the effects of shocks in the vicinity of the steady state, which we normalize to be $(A_1, \dots, A_N) = (1, \dots, 1)$. All the relevant derivatives are evaluated at that point.

THEOREM 1—Hulten (1978): *The first-order macroeconomic impact of microeconomic shocks is given by*¹⁴

$$\frac{d \log Y}{d \log A_i} = \lambda_i, \quad (1)$$

where $\lambda_i = p_i y_i / (\sum_{j=1}^N p_j c_j)$ the sales of producer i as a fraction of GDP or Domar weight.

¹⁰Our formulas can also in principle be applied with increasing returns to scale under the joint assumption of marginal-cost pricing and impossibility of shutting down production, by simply adding producer-specific fixed factors with negative marginal products and negative payments (these factors are “bads” that cannot be freely disposed of).

¹¹Shocks to the composition of demand can be captured in the same way via a set of consumer-specific productivity shocks. For example, if the final demand aggregator is CES with an elasticity strictly greater than 1, an increase in consumer demand for i can be modeled as a positive consumer-specific productivity shock to i and a set of negative consumer-specific productivity shocks to all other final goods such that the consumption-share-weighted sum of the shocks is equal to zero. The sign of the shocks must be reversed if the elasticity of substitution is strictly lower than 1, and the Cobb–Douglas case can be treated as a limit. These constructions generalize beyond the CES case. Hulten’s theorem implies that shocks to the composition of demand have no first-order effect on aggregate output, but in general, they have nonzero second-order (and more generally nonlinear) effects.

¹²To satisfactorily capture such features, one probably needs to go beyond the nested-CES case of Section 4 and use instead the nonparametric generalization to arbitrary economies provided in Section 5.

¹³If we apply the model to different periods of time and states of nature, then Y corresponds to an intertemporal aggregate consumption index reflecting intertemporal welfare.

¹⁴In the special case where A_i is a factor-augmenting shock, the relevant λ_i corresponds to a producer’s bill for this factor as a share of GDP. This is because if we relabel the labor input of producer i as a new producer, we can represent a factor-augmenting shock to i ’s labor as a Hicks-neutral shock to this new producer.

Hulten's theorem can be seen as a consequence of the first welfare theorem: since this economy is efficient, $Y(A_1, \dots, A_N)$ is also the social planning optimum and prices are the multipliers on the resource constraints for the different goods. Applying the envelope theorem to the social planning problem delivers the result.

Hulten's theorem has the powerful implication that, to a first order, the underlying microeconomic details of the structural model are completely irrelevant as long as we observe the equilibrium sales distribution: the shape of the production network, the microeconomic elasticities of substitution in production, the degree of returns to scale, and the extent to which inputs and factors can be reallocated, are all irrelevant.

We now provide a characterization of the second-order effects in terms of reduced-form elasticities. We need to introduce two objects: GE elasticities of substitution, and the input–output multiplier. Later on, we show how these reduced-form elasticities arise from structural primitives using a structural model.

We start by introducing the GE elasticities of substitution. Recall that for any homogeneous of degree 1 function $f(A_1, \dots, A_N)$, the [Morishima \(1967\)](#) elasticity of substitution is

$$\frac{1}{\rho_{ji}} = \frac{d \log(MRS_{ji})}{d \log(A_i/A_j)} = \frac{d \log(f_j/f_i)}{d \log(A_i/A_j)},$$

where MRS_{ij} is the ratio of partial derivatives with respect to A_i and A_j , and $f_i = df/dA_i$.¹⁵

When the homothetic function f corresponds to a CES utility function and A_i to quantities, ρ_{ij} is the associated elasticity of substitution parameter. However, we do not impose this interpretation, and instead treat this object as a reduced-form measure of the curvature of isoquants. By analogy, we define a pseudo elasticity of substitution for non-homothetic functions in a similar fashion.

DEFINITION 1: For any smooth function $f: \mathbb{R}^N \rightarrow \mathbb{R}$, the *pseudo elasticity of substitution* is

$$\frac{1}{\rho_{ji}} \equiv \frac{d \log(MRS_{ji})}{d \log A_i} = \frac{d \log(f_j/f_i)}{d \log A_i}.$$

The pseudo elasticity of substitution is a generalization of the Moroshima elasticity of substitution in the sense that whenever f is homogeneous of degree 1, the pseudo elasticity is the same as the Moroshima elasticity of substitution.

When applied to the equilibrium aggregate output function of a general equilibrium economy, we call the pseudo elasticity of substitution the *general equilibrium pseudo elasticity of substitution* or *GE elasticity of substitution* for short. The GE elasticity of substitution ρ_{ji} is interesting because it measures changes in the relative sales shares of j and i when there is an exogenous shock to i . This follows from the fact that

$$\begin{aligned} \frac{d \log(\lambda_i/\lambda_j)}{d \log A_i} &= \frac{d \log[(Y_i A_i)/(Y_j A_j)]}{d \log A_i} \\ &= 1 + \frac{d \log(Y_i/Y_j)}{d \log A_i} = 1 - \frac{1}{\rho_{ji}}, \end{aligned}$$

¹⁵This is a generalization of the two-variable elasticity of substitution introduced by [Hicks \(1932\)](#) and analyzed in detail by [Blackorby and Russell \(1989\)](#).

where the first equality applies Hulten's theorem. A decrease in the productivity of i causes λ_i/λ_j to increase when $\rho_{ji} \in (0, 1)$, and to decrease otherwise. We say that a j is a *GE-complement* for i if $\rho_{ji} \in (0, 1)$, and a *GE-substitute* otherwise. When f is a CES aggregator, this coincides with the standard definition of gross complements and substitutes. As usual, when f is Cobb–Douglas, i and j are neither substitutes nor complements. In general, GE-substitutability is not reflexive.

An important special case is when the shock $d \log A_f$ hits the stock of a factor. In that case, Hulten's theorem implies that $d \log Y / d \log A_f = \Lambda_f$, where $\Lambda_f = w_f \bar{l}_f / (\sum_{j=1}^N p_j c_j)$ is the share of factor f in GDP. Since $\sum_{f=1}^F w_f \bar{l}_f = \text{GDP}$, Euler's theorem implies that the aggregate output is homogeneous of degree 1 in the supplies of the factors. This implies that the general equilibrium pseudo elasticity of substitution between two factors can be interpreted as a genuine elasticity of substitution between these factors in general equilibrium.¹⁶

Next, we introduce the input–output multiplier.

DEFINITION 2: The *input–output multiplier* is

$$\xi \equiv \sum_{i=1}^N \frac{d \log Y}{d \log A_i} = \sum_{i=1}^N \lambda_i.$$

When $\xi > 1$, total sales of the shocked quantities exceed total income, an indication that there are intermediate inputs. When $\xi > 1$, the impact of a uniform technology shock is correspondingly amplified due to the fact that goods are reproducible. The input–output multiplier ξ captures the percentage change in aggregate output in response to a uniform one-percent increase in technology. Loosely speaking, it captures a notion of returns-to-scale at the aggregate level. Changes $d \log \xi / d \log A_i$ in the input–output multiplier can be interpreted as another kind of GE elasticity of substitution, namely, the substitution between the underlying factors (whose payments are GDP) and the reproducible goods (whose payments are sales).¹⁷

Having defined the GE elasticities of substitution and the input–output multiplier, we are in a position to characterize the second-order terms. We start by investigating the impact of an idiosyncratic shock.

¹⁶The difference between an elasticity of substitution and a pseudo elasticity of substitution is that the former is the elasticity of the ratio of marginal rates of substitution with respect to the ratio of two arguments, whereas the latter is the elasticity of marginal rates of substitution with respect to an argument. The two definitions are equivalent whenever the function they are applied to is homogeneous of degree 1.

¹⁷The input–output multiplier is called the intermediate input multiplier in a stylized model by Jones (2011), but it also appears under other names in many other contexts. It is also related to the network influence measure of Acemoglu et al. (2012), the granular multiplier of Gabaix (2011), the international fragmentation measure of Feenstra and Hanson (1996), the production chain length multiplier in Kim, Shin et al. (2013), and even the capital multiplier in the neoclassical growth model since capital can be treated as an intertemporal intermediate input. It also factors into how the introduction of intermediate inputs amplifies the gains from trade in Costinot and Rodriguez-Clare (2014). Although these papers feature multiplier effects due to the presence of round-about production (either via intermediate inputs or capital), they do not take into account the fact that this multiplier effect can respond to shocks. This is either because they assume Cobb–Douglas functional forms or because they focus on first-order effects.

Idiosyncratic Shocks

THEOREM 2—Second-Order Macroeconomic Impact of Microeconomic Shocks: *The second-order macroeconomic impact of microeconomic shocks is given by*¹⁸

$$\frac{d^2 \log Y}{d \log A_i^2} = \frac{d \lambda_i}{d \log A_i} = \frac{\lambda_i}{\xi} \sum_{\substack{1 \leq j \leq N \\ j \neq i}} \lambda_j \left(1 - \frac{1}{\rho_{ji}} \right) + \lambda_i \frac{d \log \xi}{d \log A_i}. \quad (2)$$

The second-order impact of a shock to i is equal to the change in i 's sales share λ_i . The change in i 's share of sales is the change in the aggregate sales to GDP ratio, minus the change in the share of sales of all other industries. The former is measured by the elasticity of the input–output multiplier ξ , while the latter depends on the GE elasticities of substitution. Collectively, the sales shares λ_i , the reduced-form elasticities ρ_{ji} , and the reduced-form elasticity of the input–output multiplier $d \log \xi / d \log A_i$ are sufficient statistics for the response of aggregate output to productivity shocks up to a second order.

This result implies that Hulten's first-order approximation is globally accurate if reduced-form elasticities are unitary $\rho_{ji} = 1$ for every j and if the input–output multiplier ξ is independent of the shock A_i . We shall see that this amounts to assuming Cobb–Douglas production and consumption functions where sales shares, and more generally the whole input–output matrix, are constant. The model is then log-linear.

Outside of this special case, there are nonlinearities, and the quality of the first-order approximation deteriorates as the shocks become bigger. The deterioration can be extreme, with the aggregate output function becoming nearly non-smooth, when ρ_{ji} approaches 0 for any j , either from above or from below. As we shall see, these arise in the cases of extreme microeconomic complementarities with no reallocation or extreme microeconomic substitutabilities with full reallocation. In these limiting cases, the first-order approximation is completely uninformative, even for arbitrarily small shocks. Similar observations apply when $d \log \xi / d \log A_i$ approaches infinity. Therefore, although the Cobb–Douglas special case is very popular in the literature, it constitutes a very special case where the second-order terms are all identically zero.¹⁹

The second-order approximation

$$\log Y \approx \log \bar{Y} + \frac{d \log Y}{d \log A_i} \log A_i + \frac{1}{2} \frac{d^2 \log Y}{d \log A_i^2} (\log A_i)^2$$

¹⁸In the case where shocks are factor-augmenting, the aggregate output function is homogeneous of degree 1 and the formula becomes

$$\frac{d^2 \log Y}{d \log A_{if}^2} = \lambda_{if} \sum_{\substack{1 \leq j \leq N \\ 1 \leq g \leq F \\ (j,g) \neq (i,f)}} \lambda_{jg} \left(1 - \frac{1}{\rho_{jg,if}} \right),$$

where A_{if} is a shock-augmenting factor f in the production of good i , λ_{if} is the share of factor f in GDP arising from its use by producer i , and $\rho_{jg,if}$ is the GE elasticity of substitution between factor f in the production of good i and factor g in the production of good j .

¹⁹See, for example, Acemoglu et al. (2012), Long and Plosser (1983), Bigio and La'O (2016), Acemoglu, Ozdaglar, and Tahbaz-Salehi (2017), Bartelme and Gorodnichenko (2015).

of the aggregate output function with respect to the productivity of producer i can then be written as

$$\log Y \approx \log \bar{Y} + \lambda_i \log A_i + \frac{1}{2} \frac{\lambda_i}{\xi} \sum_{\substack{1 \leq j \leq N \\ j \neq i}} \lambda_j \left(1 - \frac{1}{\rho_{ji}}\right) (\log A_i)^2 + \frac{1}{2} \lambda_i \frac{d \log \xi}{d \log A_i} (\log A_i)^2,$$

where \bar{Y} is Y evaluated at the steady-state technology values. When goods are GE-complements, the second-order terms amplify the effect of negative shocks and attenuate the effect of positive shocks relative to the first-order approximation. Instead, when goods are GE-substitutes, the second-order approximation attenuates the negative shocks and amplifies the positive shocks instead. A similar intuition holds for the input–output multiplier: if the input–output multiplier is increasing, then the second-order approximation amplifies positive shocks and dampens negative shocks, and if this multiplier is decreasing, then the opposite is true.

Correlated Shocks

To compute the second-order approximation

$$\begin{aligned} \log Y \approx & \log \bar{Y} + \sum_{1 \leq i \leq N} \frac{d \log Y}{d \log A_i} \log A_i + \sum_{1 \leq i \leq N} \frac{1}{2} \frac{d^2 \log Y}{d \log A_i^2} (\log A_i)^2 \\ & + \frac{1}{2} \sum_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N \\ i \neq j}} \frac{d^2 \log Y}{d \log A_j d \log A_i} \log A_i \log A_j \end{aligned}$$

of the aggregate production function with respect to shocks to several producers at once, we must extend these results to cover the off-diagonal terms in its Hessian.

PROPOSITION 3—Correlated Shocks: *The second-order macroeconomic impact of correlated microeconomic shocks is given by*

$$\begin{aligned} \frac{d^2 \log Y}{d \log A_j d \log A_i} &= \frac{d \lambda_i}{d \log A_j} \\ &= \frac{\lambda_i}{\xi} \sum_{\substack{1 \leq k \leq N \\ k \neq j}} \lambda_k \left(1 - \frac{1}{\rho_{kj}}\right) + \lambda_i \frac{d \log \xi}{d \log A_j} \\ &\quad - \lambda_i \left(1 - \frac{1}{\rho_{ij}}\right) \quad (i \neq j). \end{aligned} \quad (3)$$

The second-order effect of a common shock to i and j is not simply the sum of the second-order impacts of the idiosyncratic shocks to i and j , and instead there are interactions between the two shocks.²⁰ In Section 4, we provide an explicit characterization of the Hessian in terms of microeconomic primitives.

²⁰We can also use these ideas to capture the impact of an aggregate shock to the economy, since an aggregate shock is simply a common shock that affects all industries. If A is an aggregate productivity shock, then $\frac{d^2 \log Y}{d \log A^2} =$

Macro Moments

We can use the second-order terms to approximate an economy's macroeconomic moments. To illustrate this intuition while preserving expositional simplicity, we consider shocks to a single producer i which are log-normal with mean $\log 0$ and variance σ^2 .²¹

We first consider average log aggregate output μ_Y , for which a Taylor approximation yields

$$\mu_Y = E(\log(Y/\bar{Y})) \approx \frac{1}{2} \frac{d^2 \log Y}{d \log A_i^2} \sigma^2 = \frac{1}{2} \left(\frac{\lambda_i}{\xi} \sum_{\substack{1 \leq j \leq N \\ j \neq i}} \lambda_j \left(1 - \frac{1}{\rho_{ji}} \right) + \lambda_i \frac{d \log \xi}{d \log A_i} \right) \sigma^2.$$

Average log aggregate output can be nonzero even though the technology shocks have zero log average, and it has the same sign as the second-order term. For example, when the second-order term is negative, corresponding to GE-complementarities, average log aggregate output is lower than its deterministic steady state because nonlinearities magnify negative shocks and attenuate positive shocks.

Second-order terms also shape higher moments of the distribution of aggregate output. The variance is given by

$$\sigma_Y^2 = \text{Var}(\log(Y/\bar{Y})) \approx \left(\lambda_i^2 + 2 \left(\frac{\mu_Y}{\sigma} \right)^2 \right) \sigma^2 \geq \lambda_i^2 \sigma^2,$$

where the right-hand side is the variance of the log-linear approximation. This shows that nonlinearities tend to increase the implied variance of aggregate output.

Similarly, the skewness is given by

$$E \left(\left(\frac{\log(Y/\bar{Y}) - \mu_Y}{\sigma_Y} \right)^3 \right) \approx \frac{2\mu_Y}{\sigma_Y^3} (4\mu_Y + 3\lambda_i^2 \sigma^2).$$

For example, when the second-order term is negative, the distribution of aggregate output is skewed to the left since nonlinearities magnify negative shocks and attenuate positive shocks, even though the technology shocks are symmetric. Typical negative deviations of aggregated output are then larger than typical positive deviations.

Finally, the kurtosis is given by

$$E \left(\left(\frac{\log(Y/\bar{Y}) - \mu_Y}{\sigma_Y} \right)^4 \right) \approx 3 \left(1 + \left(\frac{\mu_Y}{\sigma} \right)^2 \frac{22(\mu_Y/\sigma)^2 + 7\lambda_i^2}{(\lambda_i^2 + 2(\mu_Y/\sigma)^2)^2} \right) \geq 3.$$

Aggregate output has excess kurtosis when second-order terms are nonzero. For example, when the second-order term is negative, the left tail is fattened because negative shocks are magnified, and this gives rise to excess kurtosis, even though the technology shocks are symmetric and thin-tailed. A higher share of the variance is then due to negative, infrequent, extreme deviations, as opposed to symmetric, frequent, and modestly sized deviations.

²¹ $\xi \sum_{i=1}^N \frac{d \log \xi}{d \log A_i}$. So, for aggregate shocks, deviations from Hulten's theorem can only come from the input-output multiplier.

²¹In Appendix E of the Supplemental Material (Baqaee and Farhi (2019)), we include the approximation equations for mean, variance, and skewness for multivariate shocks.

The importance of all these effects increases with the variance σ^2 of the shocks because they are driven by nonlinearities, and because the importance of nonlinearities increases with the size of the shocks.

Welfare Costs of Sectoral Shocks

For the majority of the paper, we focus on log aggregate output, which can be characterized with unitless elasticities. With complementarities, we have argued that sector shocks lower the mean of log aggregate output, an effect which we can interpret as the welfare cost of sectoral shocks. One may imagine that the losses from uncertainty that we identify depend on the concavity of the log function. A consumer with log utility prefers a mean-preserving reduction in uncertainty even when the aggregate output function is linear. However, as shown by Lucas (1987), the corresponding losses are extremely small in practice in business-cycle settings. The much larger effects that we identify originate in nonlinearities in production, and they are present even when the utility function is linear in aggregate consumption.

The following proposition formalizes this intuition and shows that the Lucas-style welfare losses due to nonlinearities in the utility function in the form of risk aversion and the losses due to nonlinearities in production do not interact with one another up to a second-order approximation.²²

PROPOSITION 4—Welfare Cost of Sectoral Shocks: *Let $u : \mathbb{R} \rightarrow \mathbb{R}$ be a utility function and let $Y : \mathbb{R}^N \rightarrow \mathbb{R}$ be the aggregate output function. Suppose that log productivity shocks have mean 0 and a diagonal covariance matrix with k th diagonal element σ_k^2 . Then*

$$\frac{1}{u'(\bar{Y})\bar{Y}}(E(u(Y)) - u(\bar{Y})) = -\frac{1}{2}\left(\gamma \sum_k \lambda_k^2 \sigma_k^2 - \frac{1}{\bar{Y}} \sum_k \frac{d^2 Y}{d \log A_k^2} \sigma_k^2\right), \quad (4)$$

where γ is the coefficient of relative risk aversion at the deterministic steady state \bar{Y} .

The first term on the right-hand side, which is quantitatively small, is the traditional Lucas-style cost arising from curvature in the utility function. The second term, which is quantitatively large, is due to the curvature inherent in production and does not depend on the coefficient of relative risk aversion.²³

Mapping From Micro to Macro

Theorem 2 implies that the GE elasticities of substitution ρ_{ij} and the elasticity of the input–output multiplier $d \log \xi / d \log A_i$ are sufficient statistics for the second-order impact of shocks. However, these sufficient statistics are reduced-form elasticities, and unlike λ_i and ξ , they are not readily observable. Furthermore, since they are general equilibrium objects, they cannot be identified through exogenous microeconomic variation. So,

²²In fact, it could easily be the case that a risk-averse household prefers the economy to be subject to stochastic shocks if the economy features macro-substitutability and the second-order terms are positive, which happens in the presence of GE-substitutability.

²³Proposition 4 is stated idiosyncratic shocks for expositional clarity. In Appendix A, we prove the result for more general utility functions and shocks. For our theoretical results, we find it convenient to work with elasticities $d^2 \log Y / d \log A_k^2$, but we can use these results to compute the welfare cost in Proposition 4 by noting that $d^2 Y / d A_k^2 = \bar{Y} d^2 \log Y / d \log A_k^2 - \bar{Y} \lambda_k (1 - \lambda_k)$.

while careful empirical work can identify micro-elasticities, the leap from micro-estimates to macro-effects can be hazardous.

In this paper, we provide the mapping from structural micro parameters to the reduced-form GE elasticities. This general characterization can be found in Section 4 for general nested-CES economies, and in Section 5 for arbitrary economies. However, rather than stating these results up front, we build up to the general characterization using some important special cases in Section 3.

3. ILLUSTRATIVE EXAMPLES

In this section, we work through two special cases to illustrate and isolate some intuition for how the GE elasticities of substitution and the input–output multiplier affect the shape of the aggregate output function. After working through these examples, we provide a generic characterization of the second-order terms in Sections 4 and 5.

3.1. GE Elasticities of Substitution

To start with, we focus on the GE elasticities of substitution by considering a simple example of a *horizontal economy* with no intermediate inputs. The input–output multiplier is constant and equal to 1, and so deviations from Hulten’s theorem are only due to non-unitary GE elasticities of substitution. We emphasize how the GE elasticities of substitution depend not only on the micro-elasticities of substitution, but also on the degree to which labor can be reallocated across uses and on the returns to scale in production. Throughout all the upcoming examples, variables with overlines denote steady-state values.

There are N goods produced using the production functions

$$\frac{y_i}{\bar{y}_i} = A_i \left(\frac{l_{is_i}}{\bar{l}_{is_i}} \right)^{1-\omega_g} \left(\frac{l_{ig}}{\bar{l}_{ig}} \right)^{\omega_g},$$

where l_{is_i} and l_{ig} are the amounts of the specific and general labor used by producer i . The specific labor of type i can only be used by producer i and the general labor can be used by all producers.

The household’s consumption function is

$$\frac{Y}{\bar{Y}} = \left(\sum_{i=1}^N \omega_{0i} \left(\frac{c_i}{\bar{c}_i} \right)^{\frac{\theta_0-1}{\theta_0}} \right)^{\frac{\theta_0}{\theta_0-1}},$$

where θ_0 is the microeconomic elasticity of substitution in consumption and $\sum_{i=1}^N \omega_{0i} = 1$.

The specific labors and the general labor are in fixed supplies at $\bar{l}_{s_i} = \bar{l}_{is_i}$ and $\bar{l}_g = \sum_{i=1}^N \bar{l}_{ig}$. The market-clearing conditions are

$$c_i = y_i, \quad \bar{l}_{s_i} = l_{is_i}, \quad \text{and} \quad \bar{l}_g = \sum_{i=1}^N l_{ig}.$$

Different degrees of labor reallocation can be expected depending on the degree of aggregation. The time horizon is also important since we might expect labor to be more

difficult to adjust at short horizons than at long horizons. Some of these dynamic effects can be captured by performing comparative statics with respect to ω_g , where $\omega_g = 0$ represents an economy where labor cannot be reallocated, and $\omega_g = 1$ an economy where labor can be fully reallocated.

PROPOSITION 5: *In the horizontal economy, the sales shares are given by $\lambda_i = \omega_{0i}$, the input–output multiplier is constant with $\xi = 1$ and $d \log \xi / d \log A_i = 0$. The GE elasticities of substitution are all equal and are given by*

$$\rho_{ji} = \rho = \frac{\theta_0(1 - \omega_g) + \omega_g}{\theta_0(1 - \omega_g) + \omega_g + (1 - \theta_0)}.$$

The second-order macroeconomic impact of microeconomic shocks is given by

$$\frac{d^2 \log Y}{d \log A_i^2} = \frac{d \lambda_i}{d \log A_i} = \lambda_i(1 - \lambda_i) \left(1 - \frac{1}{\rho}\right).$$

To build intuition, we consider the polar cases with no reallocation where $\omega_g = 0$ and with full reallocation where $\omega_g = 1$. We start with the no-reallocation case where $\omega_g = 0$. Because labor cannot be moved across producers, it is as if there were a fixed endowment of each good and so aggregate output is given by

$$\frac{Y}{\bar{Y}} = \left(\sum_{i=1}^N \omega_{0i} A_i^{\frac{\theta_0-1}{\theta_0}} \right)^{\frac{\theta_0}{\theta_0-1}}.$$

The GE elasticities of substitution and the microeconomic elasticities of substitution coincide so that $\rho = \theta_0$. Since $\theta_0 \in [0, \infty)$, we have $\rho \in [0, \infty)$. The second-order macroeconomic impact of microeconomic shocks is given by

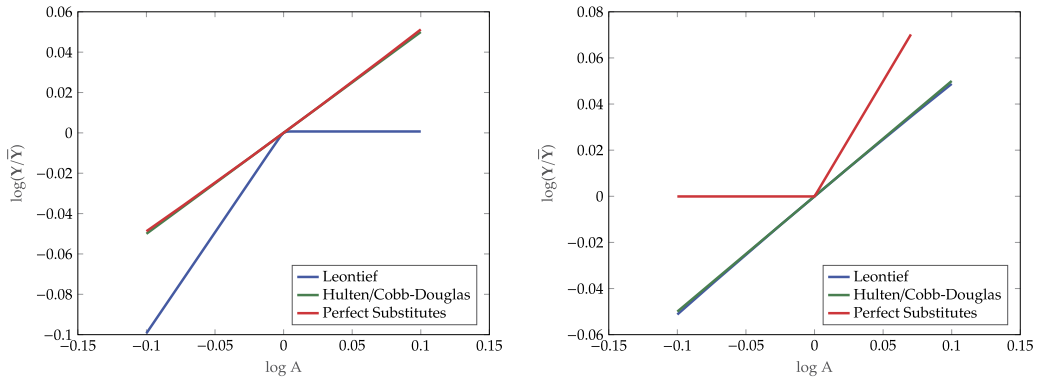
$$\frac{d^2 \log Y}{d \log A_i^2} = \frac{d \lambda_i}{d \log A_i} = \lambda_i(1 - \lambda_i) \left(1 - \frac{1}{\theta_0}\right).$$

In the Cobb–Douglas case $\theta_0 = 1$, second-order terms are identically equal to zero and the first-order approximation of Hulten’s theorem is globally accurate. The quality of the first-order approximation deteriorates as we move away from $\theta_0 = 1$ in both directions. The second-order term is negative when $\theta_0 < 1$ and positive when $\theta_0 > 1$. Relative to the first-order approximation, the second-order approximation amplifies negative shocks and mitigates positive shocks in the former case and the reverse the latter case.

To build intuition, it is useful to inspect how the relative sales share λ_i/λ_j of i versus j changes in response to a shock to i :

$$\frac{d \log(\lambda_i/\lambda_j)}{d \log A_i} = \frac{d \log(p_i/p_j)}{d \log A_i} + \frac{d \log(y_i/y_j)}{d \log A_i} = \frac{d \log(p_i/p_j)}{d \log A_i} + \frac{d \log(A_i/A_j)}{d \log A_i} = 1 - \frac{1}{\theta_0}.$$

Because there is no reallocation, the relative quantity y_i/y_j moves one-for-one with the shock to i . In the Cobb–Douglas case $\theta_0 = 1$, the relative price p_i/p_j moves one-for-one in the opposite direction, and so the relative share λ_i/λ_j remains constant. When $\theta_0 < 1$, the relative price moves more than one-for-one with the shock, and so the relative share increases when the shock is negative, and increases when it is positive. When $\theta_0 > 1$,



(a) log aggregate output with no reallocation/extreme decreasing returns. Perfect substitutes and Hulten's approximation overlap almost perfectly.

(b) log aggregate output with full reallocation/constant returns. Leontief and Hulten's approximation overlap almost perfectly.

FIGURE 1.—log aggregate output as a function of productivity $\log A_i$ in the economy with full reallocation/constant returns for different values of θ_0 . This example consists of two equally sized industries using labor as their only input. The economies depicted in Figures 1a and 1b are all equivalent to a first order.

the relative price moves less than one-for-one with the shock, and so the relative share decreases when the shock is negative, and increases when it is positive.

Consider the Leontief limit $\theta_0 \rightarrow 0$. In this limit, deviations from the first-order approximation become so large that the first-order term becomes completely uninformative. Following a negative shock to i , the relative price p_i/p_j jumps to infinity, and so does the relative share λ_i/λ_j . Following a positive shock, the relative price jumps to zero and so does the relative share. The associated amplification of negative shocks and mitigation of positive shocks is extreme.

Let us now consider the perfect-substitutes limit $\theta_0 \rightarrow \infty$. Positive shocks are amplified and negative shocks are mitigated, but the effect is not nearly so dramatic. In fact, because goods are perfect substitutes, the relative price p_i/p_j is constant. Therefore, the relative share λ_i/λ_j moves one-for-one with the shock to i . The situation is depicted graphically in Figure 1a.

Having analyzed the case with no labor reallocation, we now consider the polar opposite case, where labor can be costlessly reallocated across producers and can be used with constant returns to scale so that $\omega_g = 1$. Solving out the allocation of labor to each producer and replacing leads to the following expression for aggregate output:

$$\frac{Y}{\bar{Y}} = \frac{\left(\sum_{i=1}^N \omega_{0i} A_i^{\theta_0-1} \right)^{\frac{\theta_0}{\theta_0-1}}}{\sum_{i=1}^N \omega_{0i} A_i^{\theta_0-1}}.$$

In this case, the GE elasticities of substitution do *not* typically coincide with the structural microeconomic elasticity of substitution since we have $\rho = 1/(2 - \theta_0)$. Because $\theta_0 \in [0, \infty)$, we have $\rho \in (-\infty, 0) \cup [1/2, +\infty)$. The second-order macroeconomic im-

pact of microeconomic shocks is given by

$$\frac{d^2 \log Y}{d \log A_i^2} = \frac{d \lambda_i}{d \log A_i} = \lambda_i(1 - \lambda_i)(\theta_0 - 1).$$

As above, in the Cobb–Douglas case $\theta_0 = 1$, second-order terms are identically equal to zero and the first-order approximation of Hulten’s theorem is globally accurate. The second-order term is negative when $\theta_0 < 1$ and positive when $\theta_0 > 1$. Relative to the first-order approximation, the second-order approximation amplifies negative shocks and mitigates positive shocks in the former case and the reverse in the latter case. However, this time, the second-order term becomes singular when the goods are highly substitutable rather than when they are highly complementary.

Once again, we can unpack this result by noting that

$$\frac{d \log(\lambda_i/\lambda_j)}{d \log A_i} = \frac{d \log(p_i/p_j)}{d \log A_i} + \frac{d \log(y_i/y_j)}{d \log A_i} = \frac{d \log(A_j/A_i)}{d \log A_i} + \frac{d \log(y_i/y_j)}{d \log A_i} = \theta_0 - 1.$$

Because labor can be costlessly reallocated across producers, the relative price p_i/p_j always moves inversely one-for-one with the shock to i . In the Cobb–Douglas case, the relative quantity y_i/y_j moves one-for-one with the shock to i , and the relative share λ_i/λ_j remains constant. When $\theta_0 < 1$, the relative quantity moves less than one-for-one with the shock as labor is reallocated toward i if the shock is negative and away from i if the shock is positive. As a result, the relative share increases when the shock is negative, and increases when it is positive. When $\theta_0 > 1$, the relative quantity moves more than one-for-one with the shock as labor is reallocated away from i when the shock is negative and toward i when it is positive. As a result, the relative share decreases when the shock is negative, and increases when it is positive.

Contrary to what one may have assumed, a near-Leontief production function is *not* sufficient for generating large deviations from Hulten’s theorem, as long as factors can be reallocated freely, precisely because this reallocation is successful at reinforcing “weak links.” In the Leontief limit, the relative quantity y_i/y_j is invariant to the shock, and so the relative sales share λ_i/λ_j moves inversely one-for-one with the shock to i . Relative to the first-order approximation, the second-order approximation still amplifies negative shocks and mitigates positive shocks, but the corresponding magnitudes are much smaller than in the case where labor cannot be reallocated.

In the perfect-substitutes limit, labor is entirely allocated to the most productive producer. In response to a positive shock to i , the relative quantity y_i/y_j jumps to infinity, and so does the relative share λ_i/λ_j . In response to a negative shock, the relative quantity drops to zero, and so does the relative share. Relative to the first-order approximation, the second-order approximation still amplifies positive shocks and mitigates negative shocks, but the corresponding magnitudes are now much larger than in the case where labor cannot be reallocated. The situation is depicted graphically in Figure 1b.

Finally, note that both when labor can or cannot be reallocated, the second-order term scales in $\lambda_i(1 - \lambda_i)$ as a function of the size λ_i of the shocked producer i . Its absolute value is therefore hump-shaped in λ_i : it goes to zero when λ_i is close to 0 or 1, and reaches a maximum when λ_i is intermediate at $1/2$. That the term is small when λ_i is close to 0 is intuitive. That it is small when λ_i is close to 1 makes sense since then the economy behaves much like producer i and aggregate output is then close to being proportional to A_i . The second-order term can only be significant for intermediate values of λ_i .

To recap, with complementarities: a negative shock can cause a large downturn when labor cannot be freely reallocated, but the ability to reallocate labor largely mitigates these effects; positive shocks have a lesser impact. By contrast, with substitutabilities: a positive shock can cause a big boom when labor can be reallocated, but the inability to reallocate labor mitigates these effects; negative shocks have a lesser impact. Cobb–Douglas stands as a special case where the macroeconomic impact of microeconomic shocks is symmetric independently of whether or not labor can be reallocated (since the equilibrium allocation of labor across producers is constant even when labor can be reallocated).²⁴ These effects are less pronounced when the size of the shocked producer is very small or very large, and are more pronounced when it is intermediate.

3.2. Input–Output Multiplier

In the previous example of a horizontal economy, the input–output multiplier ξ is constant and deviations from Hulten’s theorem are due to non-unitary GE elasticities of substitution. We now focus on a different example, that of a *roundabout economy*, where deviations from Hulten’s theorem are driven purely by variability in ξ , and the GE elasticities of substitution play no role.

The economy has a single good and single factor. Gross output is given by

$$\frac{y_1}{\bar{y}_1} = A_1 \left(\omega_{1l} \left(\frac{l_1}{\bar{l}_1} \right)^{\frac{\theta_1-1}{\theta_1}} + (1 - \omega_{1l}) \left(\frac{x_1}{\bar{x}_1} \right)^{\frac{\theta_1-1}{\theta_1}} \right)^{\frac{\theta_1}{\theta_1-1}},$$

where x_1 is the amount of good 1 used as an intermediate input. The supply of the factor is inelastic at $\bar{l} = \bar{l}_1$. Final output $Y = c_1$ is produced one-to-one from good 1.

The market-clearing conditions are

$$y_1 = c_1 + x_1 \quad \text{and} \quad \bar{l} = l_1.$$

The steady-state input–output multiplier

$$\xi = 1 + (1 - \omega_{1l}) + (1 - \omega_{1l})^2 + \cdots = 1/\omega_{1l}$$

decreases with the labor share ω_{1l} and increases with the intermediate input share $1 - \omega_{1l}$. Hulten’s theorem implies that

$$\frac{d \log Y}{d \log A_1} = \xi,$$

so that the first-order impact of the shock increases with the steady-state input–output multiplier ξ .

²⁴These results are closely related to the findings in Jones (2011), who noted that the relevant CES parameter used in aggregating microeconomic productivity shocks depends on whether or not factors are allocated through the market or assigned exogenously.

PROPOSITION 6—Variable IO Multiplier: *In the roundabout economy, the input–output multiplier is given by $\xi = 1/\omega_{1l}$ and its elasticity is given by*²⁵

$$\frac{d \log \xi}{d \log A_1} = (\xi - 1)(\theta_1 - 1).$$

The second-order macroeconomic impact of microeconomic shocks is given by

$$\frac{d^2 \log Y}{d \log A_1^2} = \frac{d\xi}{d \log A_1} = \xi(\xi - 1)(\theta_1 - 1).$$

Hulten’s approximation is exact only when there are no intermediate inputs so that $\xi = 1$ or when the economy is Cobb–Douglas so that $\theta_1 = 1$. Otherwise, the second-order term is increasing in $\theta_1 - 1$ and in a network term $\xi(\xi - 1)$.²⁶

Intuitively, this results from the fact that output is used as its own input. When $\theta_1 = 1$, the input–output multiplier remains constant. When $\theta_1 < 1$, the input–output multiplier increases if the shock is negative, and decreases if it is positive. When $\theta_1 > 1$, the input–output multiplier decreases if the shock is negative, and increases if it is positive. The larger is the steady-state input–output multiplier, the larger is the effect.

Figure 2 plots $\log Y$ as a function of $\log A_1$ for the case where $\theta_1 \approx 0$, $\theta_1 = 1$, and $\theta_1 = 2$. In the limit $\theta_1 \rightarrow 0$, output is linear in productivity (rather than log-linear) with slope $1/\omega_{1l}$. When $\theta_1 = 2$, output is hyperbolic in productivity.²⁷

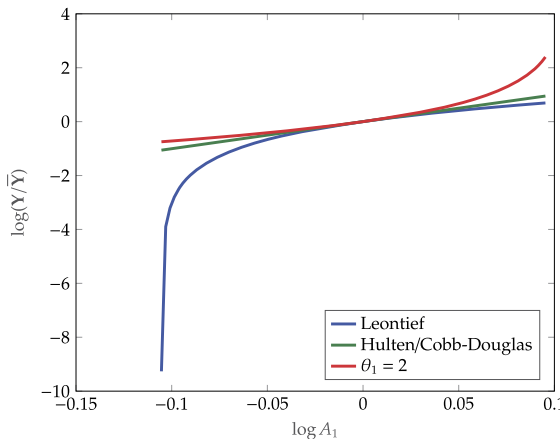


FIGURE 2.—Output as a function of productivity shocks $\log A_1$ with variable input–output multiplier effect with steady-state input–output multiplier $\xi = 10$.

²⁵ Proposition 6 shows that even though the gross production function is homogeneous in productivity, aggregate net output (value added) is not homogeneous of degree 1. Furthermore, aggregate output is not homogeneous of any degree in equilibrium, since ξ varies in response to the shock.

²⁶ There is no closed-form solution for equilibrium aggregate output in this case.

²⁷ In this example, when $\theta_1 = 0$, we have $Y = A/\omega_{1l}$, where $1/\omega_{1l}$ is the steady-state input–output multiplier. Therefore, although Hulten’s approximation fails in log terms, Hulten’s theorem is globally accurate in *linear* terms. This is a consequence of the fact that there is only one good. In Appendix B of the Supplemental Material, we generalize this example to multiple goods, and show that output can be very strongly nonlinear even with full labor reallocation.

4. GENERAL NESTED-CES NETWORKS

We now characterize the second-order terms for a general nested-CES economy (encompassing the examples in Section 3). Throughout this section, variables with over-lines are normalizing constants equal to the values in steady state.²⁸

Any nested-CES economy with a representative consumer, an arbitrary numbers of nests, elasticities, and intermediate input use, can be rewritten in what we call *standard form*, which simply means that each CES aggregator corresponds to a node in the production network with a 1 node-specific elasticity of substitution. Through a relabeling, this structure can represent any nested-CES economy with an arbitrary pattern of nests and elasticities. Intuitively, by relabeling each CES aggregator to be a new producer, we can have as many nests as desired.

Formally, a nested-CES economy in standard form is defined by a tuple (ω, θ, F) and a set of normalizing constants (\bar{y}, \bar{x}) . The $(N + 1 + F) \times (N + 1 + F)$ matrix ω is a matrix of input–output parameters where the first row and column correspond to household sector, the next N rows and columns correspond to reproducible goods, and the last F rows and columns correspond to factors. What distinguishes factors from goods is that factors cannot be produced. The $(N + 1) \times 1$ vector θ is a vector of microeconomic elasticities of substitution. For convenience, we use number indices starting at 0 instead of 1 to describe the elements of ω and θ .²⁹

The F factors are modeled as non-reproducible goods and the production function of these goods are endowments

$$\frac{y_f}{\bar{y}_f} = 1.$$

The other $N + 1$ goods are reproducible with production functions

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

where x_{ij} are intermediate inputs from j used by i . Producer 0 represents final demand and its production function is the final-demand aggregator so that

$$\frac{Y}{\bar{Y}} = \frac{y_0}{\bar{y}_0},$$

where Y is aggregate output and y_0 is the final good.

The market-clearing conditions for goods and factors $0 \leq i \leq N + F$ are

$$y_i = \sum_{j=0}^N x_{ji}.$$

To state our results, we need the following definitions.

²⁸Since we are interested in log changes, the normalizing constants are irrelevant. We use normalized quantities since it simplifies calibration, and clarifies the fact that CES aggregators are not unit-less.

²⁹We impose the restriction that $\omega_{ij} \in [0, 1]$, $\sum_{j=1}^{N+F} \omega_{ij} = 1$ for all $0 \leq i \leq N$, $\omega_{fj} = 0$ for all $N + 1 \leq f \leq N + F$, $\omega_{0f} = 0$ for all $N + 1 \leq f \leq N + F$, and $\omega_{i0} = 0$ for all $0 \leq i \leq N$.

DEFINITION 3: The $(N + 1 + F) \times (N + 1 + F)$ *input-output* matrix Ω is the matrix whose ij th element is equal to the steady-state value of

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

The *Leontief inverse* is

$$\Psi = (I - \Omega)^{-1}.$$

Intuitively, the ij th element Ψ_{ij} of the Leontief inverse is a measure of i 's total reliance on j as a supplier. It captures both the direct and indirect ways through which i uses j in its production.³⁰

DEFINITION 4: The *input-output covariance operator* is

$$\text{Cov}_{\Omega^{(k)}}(\Psi_{(i)}, \Psi_{(j)}) = \sum_{l=1}^{N+F} \Omega_{kl} \Psi_{li} \Psi_{lj} - \left(\sum_{l=1}^{N+F} \Omega_{kl} \Psi_{li} \right) \left(\sum_{l=1}^{N+F} \Omega_{kl} \Psi_{lj} \right). \quad (5)$$

It is the covariance between the i th and j th column of the Leontief inverse using the k th row of the input-output matrix as the distribution. The *input-output covariance operator* plays a crucial role in our results.

We consider arbitrary CES network structures (in standard form), starting with a single factor and then generalizing to multiple factors. As previously mentioned, a one-factor model is equivalent to a model where primary factors are equivalent and can be fully reallocated. To model limited factor reallocation or decreasing returns, we need to have multiple factors.

4.1. One Factor

PROPOSITION 7—Second-Order Network Centrality: *Consider a nested-CES model in standard form with a single factor. The second-order macroeconomic impact of microeconomic shocks is given by*

$$\frac{d^2 \log Y}{d \log A_j d \log A_i} = \frac{d \lambda_i}{d \log A_j} = \sum_{k=0}^N (\theta_k - 1) \lambda_k \text{Cov}_{\Omega^{(k)}}(\Psi_{(i)}, \Psi_{(j)}), \quad (6)$$

and in particular,

$$\frac{d^2 \log Y}{d \log A_i^2} = \frac{d \lambda_i}{d \log A_i} = \sum_{k=0}^N (\theta_k - 1) \lambda_k \text{Var}_{\Omega^{(k)}}(\Psi_{(i)}). \quad (7)$$

Equations (6) and (7) have a simple intuition. Let us focus first on equation (6). The change in the sales share of i , in response to a shock to j , depends on how the relative demand expenditure for i changes. Changes in the demand expenditure for i arise from

³⁰See, for example, Baqaee (2015) for a detailed description of Ω and Ψ .

the substitution by the different nodes k and are captured by the different terms in the sum on the right-hand side.³¹

Consider, for example, the effect of a negative productivity shock $d \log A_j < 0$ to j . The change in the vector of prices of the different producers is proportional to the vector of direct and indirect exposures to the shock, which is simply the j th column $\Psi_{(j)}$ of the Leontief inverse. Now consider a given producer k . If $\theta_k < 1$, producer k increases its expenditure share on inputs whose price increases more, that is, inputs that are more exposed to the shock to j , as measured by $\Psi_{(j)}$. This increases the relative demand expenditure for i if those inputs are also relatively more exposed to i , as measured by the i th column of the Leontief inverse $\Psi_{(i)}$. The overall effect is stronger, the higher is the covariance $\text{Cov}_{\Omega(k)}(\Psi_{(i)}, \Psi_{(j)})$, the larger is the size of producer k as measured by λ_k , and the further away from 1 is the elasticity of substitution θ_k as measured by $\theta_k - 1$.

Equation (7) is a particular case of equation (6) and so the intuition is identical. The change in the sales share of i depends on substitution by all producers k . The extent to which substitution by producer k matters depends on how unequally k is exposed to i through its different inputs, on how large k is, and on how far away from 1 is the elasticity of substitution in production of k . If k is small, or is exposed in the same way to i through all of its inputs, then the extent to which it can substitute amongst its inputs is irrelevant. If the elasticity of substitution of k is equal to 1, then the direct and indirect relative demand expenditure for i arising from k does not change in response to shocks. Equation (7) can be seen as a centrality measure which combines structural microeconomic elasticities of substitution and features of the network.³²

The Cobb–Douglas specification is the knife-edge special case where all the second-order terms are equal to zero and where the first-order approximation is globally accurate. This occurs because sales shares, and more generally, the whole input–output matrix, are constant and can be taken to be exogenous. Away from the Cobb–Douglas case, sales shares and the input–output matrix respond endogenously to shocks, and this is precisely what gives rise to the nonlinearities which are captured by the second-order approximation.

GE Elasticities of Substitution

Proposition 7 can also be used to compute the GE elasticities of substitution, using equations (6) and (7) to substitute the corresponding derivatives in the following equations:

$$1 - \frac{1}{\rho_{ji}} = \frac{d \log \lambda_i}{d \log A_i} - \frac{d \log \lambda_j}{d \log A_i}$$

and

$$\frac{d \log \xi}{d \log A_i} = \frac{1}{\xi} \sum_j \lambda_j \frac{d \log \lambda_j}{d \log A_i}.$$

³¹In ongoing work (Baqaee and Farhi (2019)), we show that there is a connection between these formulas and the gains-from-trade formulas in Arkolakis, Costinot, and Rodríguez-Clare (2012). See Appendix F of the Supplemental Material for more details.

³²Equation (7) is also related to the *concentration* centrality defined by Acemoglu, Ozdaglar, and Tahbaz-Salehi (2016), but generalizes their result by allowing for heterogeneity in the interaction functions, nonsymmetric network structures, and micro-founds its use for production networks.

A Network-Irrelevance Result

To build more intuition, we provide a benchmark irrelevance result where the deviation from Hulten's approximation does not depend on the network structure.³³ The key assumptions required for obtaining this irrelevance result are: (1) productivity shocks are factor-augmenting; (2) the structural microeconomic elasticities of substitution are all the same.

Once the economy is written in standard form, the shocks are factor-augmenting if they only hit producers i which have the primary factor as their only input, i.e. if $\Omega_{ij} = 0$ for every $j = 0, \dots, N$. In other words, the productivity shocks A_i hitting producers i which do not have the primary factor as their only input are kept at their steady-state values of $A_i = 1$.

COROLLARY 1—Network Irrelevance: *Consider a nested-CES model in standard form with a single factor, uniform elasticities of substitution $\theta_j = \theta$ for every j , and with factor-augmenting shocks. Aggregate output is given by the closed-form expression*

$$\frac{Y}{\bar{Y}} = \left(\sum_{i=0}^N \bar{\lambda}_i A_i^{\theta-1} \right)^{\frac{1}{\theta-1}},$$

where $\bar{\lambda}_i$ is the steady-state Domar weight of i . The second-order macroeconomic impact of factor-augmenting microeconomic shocks is given by

$$\frac{d^2 \log Y}{d \log A_j d \log A_i} = \frac{d \lambda_i}{d \log A_j} = (\theta - 1) \lambda_i (\mathbf{1}(i = j) - \lambda_j),$$

and in particular,

$$\frac{d^2 \log Y}{d \log A_i^2} = \frac{d \lambda_i}{d \log A_i} = \sum_{j=0}^N (\theta_j - 1) \lambda_j \text{Var}_{\Omega(j)}(\Psi_{(i)}) = (\theta - 1) \lambda_i (1 - \lambda_i).$$

In words, if we consider factor-augmenting shocks, and if all microeconomic elasticities of substitution are the same, then the network structure remains irrelevant, even though there are deviations from Hulten's approximation. In this special case, the Domar weights and the structural microeconomic elasticities of substitution are sufficient statistics for the second-order effects. In fact, the result is true not only locally, but also globally.

Essentially, factor-augmenting shocks shut down variations in ξ , which is constant and equal to 1, and uniform structural microeconomic elasticities of substitution shut down variations in ρ_{ji} which are uniform and constant

$$\rho_{ji} = \frac{1}{2 - \theta}.$$

Deviating from either condition breaks the irrelevance.

³³Footnote 13 of Baqaee (2018) also discusses this network irrelevance result.

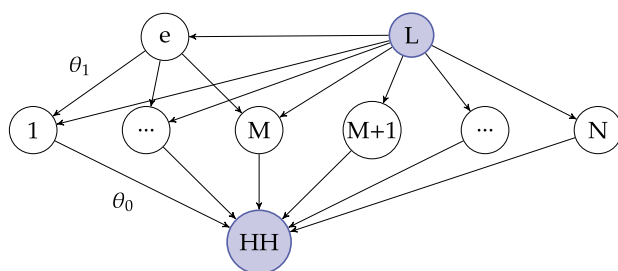


FIGURE 3.—An illustration of the economy with a near-universal intermediate input which we call energy. Each downstream producer substitutes across labor and energy with elasticity $\theta_1 < 1$. The household can substitute across final goods with elasticity of substitution $\theta_0 > \theta_1$. Energy is produced from labor with constant returns.

Energy Example—One Factor

A simple example, motivated by a universal intermediate input like energy, helps explain some of the intuition of Proposition 7 and Corollary 1. Consider the example economy depicted in Figure 3. Energy is produced linearly from labor:

$$\frac{e}{\bar{e}} = A_e \frac{l_e}{\bar{l}_e}.$$

Downstream producers produce using energy and labor with elasticity of substitution θ_1 :

$$\frac{y_i}{\bar{y}_i} = \left((1 - \omega_{ie}) \frac{l_i}{\bar{l}_i}^{\frac{\theta_1 - 1}{\theta_1}} + \omega_{ie} \left(\frac{e_i}{\bar{e}_i} \right)^{\frac{\theta_1 - 1}{\theta_1}} \right)^{\frac{\theta_1}{\theta_1 - 1}}.$$

They sell directly to the household who values goods with an elasticity of substitution θ_0 :

$$\frac{Y}{\bar{Y}} = \left(\sum_{i=1}^N \omega_{0i} \left(\frac{c_i}{\bar{c}_i} \right)^{\frac{\theta_0 - 1}{\theta_0}} \right)^{\frac{\theta_0}{\theta_0 - 1}},$$

where $\sum_{i=1}^N \omega_{0i} = 1$.

The market-clearing conditions are

$$y_i = c_i, \quad e = \sum_{i=1}^N e_i \quad \text{and} \quad \bar{l} = l_e + \sum_{i=1}^N l_i,$$

where $\bar{l} = \bar{l}_e + \sum_{i=1}^N \bar{l}_i$.

Producer i 's steady-state sales share is $\lambda_i = \omega_{0i}$, the intermediate input share of industry i is ω_{ie} , and the sales share of energy is $\lambda_e = \sum_i \lambda_i \omega_{ie}$.

We simplify the example further by supposing that all final sectors are equally sized with $\lambda_i = 1/N$, and that $M \leq N$ producers use energy with the same steady-state intermediate input share $\omega_{ie} = \omega_e$, while the other $N - M$ producers use no energy at all so that $\omega_{ie} = 0$. We set ω_e to ensure that λ_e stays constant. We take $\theta_1 < \theta_0$ and $\theta_1 < 1$.

Proposition 7 implies that

$$\begin{aligned}\frac{d^2 \log Y}{d \log A_e^2} &= \lambda_e^2 \left(\frac{N}{M} - 1 \right) (\theta_0 - 1) + \lambda_e \left(1 - \frac{N}{M} \lambda_e \right) (\theta_1 - 1) \\ &= \lambda_e (1 - \lambda_e) (\theta_0 - 1) + \lambda_e \left(1 - \frac{N}{M} \lambda_e \right) (\theta_1 - \theta_0).\end{aligned}$$

The first equation directly expresses the second-order term as a weighted sum of the microeconomic elasticities of substitution as in Proposition 7. The first term on the right-hand side of the second equation is the network-independent second-order term when all the microeconomic elasticities of substitution are identical so that $\theta_0 = \theta_1$ as in Corollary 1. The second term is a network-dependent correction that takes into account the fact that $\theta_0 \neq \theta_1$.

When every sector uses energy $M = N$, these equations become

$$\frac{d^2 \log Y}{d \log A_e^2} = \lambda_e (1 - \lambda_e) (\theta_1 - 1),$$

and the elasticity of substitution in consumption θ_0 drops out completely. The fact that θ_0 is irrelevant when $M = N$ is a manifestation of the general principle stated in Proposition 7. When $M = N$, energy is a universal input and hence $\text{Var}_{\Omega^{(0)}}(\Psi_{(e)}) = 0$. In this case, the household is symmetrically exposed to shocks to energy via the different downstream producers, and so the elasticity of substitution in consumption θ_0 is irrelevant.

When $M \neq N$ instead, θ_0 matters, with a weight that decreases with N . Through the lens of Proposition 7, $\text{Var}_{\Omega^{(0)}}(\Psi_{(e)}) = \lambda_e^2 (N/M - 1) > 0$ is decreasing in M : as heterogeneity in energy intensity across downstream producers increases, the ability of the household to substitute across these producers matters more and more.

Because $\theta_1 < 1$, when $M = N$, the second-order approximation magnifies negative the macroeconomic impact of negative shocks to energy compared to the first-order approximation. As M decreases, this effect becomes weaker, since a lower M means that energy is less of a universal input, and so it becomes easier to substitute away from it further downstream across producers with different energy intensities. The sign of the effect can even flip if M is low enough and if θ_0 is high enough above 1.

Macro-Influence—One Factor

A final implication of Proposition 7 is that it is only the producer's role as a supplier that matters, not its role as a consumer.³⁴

PROPOSITION 8—Macro-Influence: *Consider a nested-CES model in standard form with a single factor. Suppose that all producers k have the same expenditures on producers i and j so that $\Omega_{ki} = \Omega_{kj}$ for all k . Then*

$$\begin{aligned}\frac{d \log Y}{d \log A_i} &= \frac{d \log Y}{d \log A_j}, \\ \frac{d^2 \log Y}{d \log A_i^2} &= \frac{d^2 \log Y}{d \log A_j^2},\end{aligned}$$

³⁴This generalizes a result in Baqaee (2018).

and for all l ,

$$\frac{d^2 \log Y}{d \log A_l d \log A_i} = \frac{d^2 \log Y}{d \log A_l d \log A_j}.$$

The intuition is that, in a one-factor model, we can normalize the wage to 1, and then aggregate output, which is equal to real factor income, depends only on the prices of final goods. A change in the size of the i th industry does not affect its price. Hence, a productivity shock travels downstream from suppliers to their consumers by lowering their marginal costs and hence their prices, but it does not travel upstream from consumers to their suppliers. This result fails whenever there are multiple factors, and by implication when the model does not feature constant returns to scale.³⁵

4.2. Multiple Factors

We now generalize the results of the previous section to allow for multiple factors of production. This in turn opens the door to modeling limited-reallocation and decreasing-returns-to-scale via producer and industry-specific fixed factors.

We sometime use separate uppercase indices to denote the producers that correspond to factors, and lowercase indices to denote all other producers. For example, we sometime use Λ_f to denote the Domar weight, or income share, of factor f , and Λ to denote the $F \times 1$ vector of factor shares.

PROPOSITION 9—Second-Order Network Centrality With Multiple Factors: *Consider a nested-CES model in standard form. Then*

$$\begin{aligned} \frac{d^2 \log Y}{d \log A_j d \log A_i} &= \frac{d \lambda_i}{d \log A_j} \\ &= \sum_{k=0}^N (\theta_k - 1) \lambda_k \text{Cov}_{\Omega^{(k)}}(\Psi_{(j)}, \Psi_{(i)}) \\ &\quad - \sum_{f=N+1}^{N+F} \frac{d \log \Lambda_f}{d \log A_j} \sum_{k=0}^N (\theta_k - 1) \lambda_k \text{Cov}_{\Omega^{(k)}}(\Psi_{(f)}, \Psi_{(i)}), \end{aligned} \quad (8)$$

where the vector of elasticities of the factor income shares to the shocks solves the linear system

$$\frac{d \log \Lambda}{d \log A_j} = \Gamma \frac{d \log \Lambda}{d \log A_j} + \delta^{(j)}, \quad (9)$$

with

$$\Gamma_{f,g} = -\frac{1}{\Lambda_f} \left(\sum_{k=0}^N (\theta_k - 1) \lambda_k \text{Cov}_{\Omega^{(k)}}(\Psi_{(f)}, \Psi_{(g)}) \right),$$

³⁵Acemoglu, Akcigit, and Kerr (2015) showed that outside of the Cobb–Douglas special case, shocks can propagate both upstream and downstream. There is no contradiction between their result and Proposition 8. Their results and ours simply characterize different forms of propagation: they focused on the propagation of shocks to producers on the quantities and sales of other producers, whereas we focus on the impact of shocks to producers on the prices of other producers and on aggregate output.

and

$$\delta_f^{(j)} = \frac{1}{\Lambda_f} \left(\sum_{k=0}^N (\theta_k - 1) \lambda_k \text{Cov}_{\Omega^{(k)}}(\Psi_{(f)}, \Psi_{(j)}) \right).$$

Note that we can rewrite equation (8) as a function of $d \log \lambda_j / d \log A_j$ using the identity $d \log \lambda_j / d \log A_j = (1/\lambda_j)(d \lambda_j / d \log A_j)$. Proposition 9 can then be seen as a full characterization of the elasticities of the Domar weights of the different producers to the different shocks.

The intuition is the following. The first set of summands on the right-hand side of equation (8) are exactly those in equation (6) in Proposition 7: these terms capture how substitution by downstream producers k in response to a shock to j changes the sales share of i . The second set of summands in equation (8) take into account the fact that, when there are multiple factors, the shock also changes relative factor prices, and substitution in response to changes in factor prices in turn affects the sales share of i .

Consider, for example, a negative shock $d \log A_j < 0$ to producer j . Imagine that this shock increases the price of factor f relative to the prices of other factors, so that $d \log \Lambda_f > 0$. Now consider the response of a producer k to this change. If $\theta_k < 1$, producer k increases its expenditure share on producers that are more exposed to factor f as measured by $\Psi_{(f)}$. If these producers are also more exposed to i , as measured by $\Psi_{(i)}$, then the substitution increases the sales share of i . These changes must be cumulated across producers k and factors f . The total effect on the relative demand expenditure for producer i , and hence on its sales share, is the sum of the effect of substitutions in response to the initial impulse $d \log A_j$, as well as the substitutions in responses to changes in relative factor prices captured by $d \log \Lambda_f$.

Equation (9) in turn determines how factor shares $d \log \Lambda_f / d \log A_j$ respond to different shocks. For a given set of factor prices, a shock to j affects the relative demand expenditure for each factor, and hence the factor income shares, as measured by the $F \times 1$ vector $\delta^{(j)}$. This change in the factor income shares then causes further substitution through the network, leading to additional changes in relative factor shares and prices. The impact of the change in the relative share or price of factor g on the relative demand expenditure for factor f is measured by the fg th element of the $F \times F$ matrix Γ . Crucially, the matrix Γ does not depend on which producer j is being shocked.

We can verify that we get back Proposition 7 when there is only a single factor, since in that case the exposure vector $\Psi_{(f)}$, corresponding to the unique factor, is equal to a vector of all ones, and so the second set of summands in equation (8) is identically zero.

Just like in the case of a single factor and for the same reasons, the Cobb–Douglas specification is the knife-edge special case where all the second-order terms are equal to zero and where the first-order approximation is globally accurate because sales shares, and more generally the whole input–output matrix, are constant and can be taken to be exogenous.

GE Elasticities of Substitution

Proposition 9 can also be used to compute the GE elasticities of substitution, using equations (6) and (7) to substitute the corresponding derivatives in the following equa-

tions³⁶:

$$1 - \frac{1}{\rho_{ji}} = \frac{d \log \lambda_i}{d \log A_i} - \frac{d \log \lambda_j}{d \log A_i}$$

and

$$\frac{d \log \xi}{d \log A_i} = \frac{1}{\xi} \sum_j \lambda_j \frac{d \log \lambda_j}{d \log A_i}.$$

A Network-Irrelevance Result

In the special case where all microeconomic elasticities of substitution are the same, we once again obtain a network-irrelevance result. However, because there are multiple factors, it is not enough to consider factor-augmenting shocks as we did in the case of a single factor, and we must instead focus on shocks that increase the overall quantities of the different factors.

COROLLARY 2—Network Irrelevance: *Consider a nested-CES model in standard form with uniform elasticities of substitution $\theta_j = \theta$ for every j , and shocks A_f to the supplies of the different factors f . Aggregate output is given by the following closed-form expression:*

$$\frac{Y}{\bar{Y}} = \left(\sum_{f=N+1}^{N+F} \bar{\Lambda}_f A_f^{\frac{\theta-1}{\theta}} \right)^{\frac{\theta}{\theta-1}},$$

where $\bar{\Lambda}_f$ is the Domar weight of f at steady state. The second-order macroeconomic impact of microeconomic shocks to the supplies of factors is given by

$$\frac{d^2 \log Y}{d \log A_g d \log A_f} = \frac{d \Lambda_f}{d \log A_g} = \left(1 - \frac{1}{\theta} \right) \Lambda_g (\mathbf{1}(g=f) - \Lambda_g),$$

and in particular,

$$\frac{d^2 \log Y}{d \log A_f^2} = \frac{d \Lambda_f}{d \log A_f} = \left(1 - \frac{1}{\theta} \right) \Lambda_f (1 - \Lambda_f).$$

³⁶Although Proposition 9 is stated in terms of productivity shocks to non-factor producers j , the same formulas hold for a productivity shock $d \log A_g$ to a factor industry g . The productivity shock is then just a shock to the endowment of the factor. For example, we have

$$\frac{d^2 \log Y}{d \log A_g^2} = \Lambda_g \frac{d \log \Lambda_g}{d \log A_g},$$

where $d \log \Lambda_g / d \log A_g$ is the g th element of the vector $d \log \Lambda / d \log A_g$ which solves the linear system

$$\frac{d \log \Lambda}{d \log A_g} = \Gamma \frac{d \log \Lambda}{d \log A_g} + \delta^{(g)}. \quad (10)$$

We can also compute the GE elasticities of substitution between two factors g and f using

$$1 - \frac{1}{\rho_{gf}} = \frac{d \log \Lambda_f}{d \log A_f} - \frac{d \log \Lambda_g}{d \log A_f}.$$

We refer the reader to Baqaee and Farhi (2018a) for an extensive discussion of macroeconomic elasticities of substitution between factors. This paper also characterizes the macroeconomic bias of technical change as a function of microeconomic primitives.

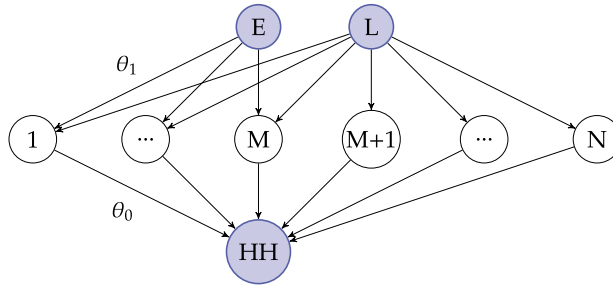


FIGURE 4.—An illustration of the economy with a near-universal input which we treat as energy. Each industry has different shares of labor and energy and substitutes across labor and energy with elasticity $\theta_1 < 1$. The household can substitute across goods with elasticity of substitution $\theta_0 > \theta_1$. Labor and energy are in fixed supply.

In this special case, the Domar weight and the structural microeconomic elasticities of substitution are sufficient statistics for the second-order effects. In fact, the result is true not only locally, but also globally.

A consequence of this corollary is that whenever all the micro-elasticities of substitution are the same and equal to θ , the GE elasticity of substitution ρ_{gf} between any two factors f and g is also equal to θ .

Energy Example—Multiple Factors

We revisit the energy example introduced in Section 4.1, but we now model energy as an endowment rather than as a produced good. The economy is represented in Figure 4. For this example, the effect of a shock to energy is now a nonlinear function of the underlying microeconomic elasticities of substitution

$$\frac{d^2 \log Y}{d \log A_e^2} = \frac{d A_e}{d \log A_e^2} = \frac{(\theta_0 - 1) \Lambda_e (1 - \Lambda_e) + (\theta_1 - \theta_0) \Lambda_e \left(1 - \frac{N}{M} \Lambda_e\right)}{\theta_0 + (\theta_1 - \theta_0) \frac{1 - \frac{N}{M} \Lambda_e}{1 - \Lambda_e}}.$$

The difference with the case of a single factor is that following a negative productivity shock to energy, labor cannot be reallocated to the production of energy in order to reinforce that weak link. This effect, encapsulated in the denominator, further amplifies the effect of the shock.

Once again, and in accordance with Corollary 2, whenever the micro-elasticities of substitution are the same $\theta_0 = \theta_1$, the shape of the network becomes irrelevant. Additionally, in the extreme case where energy becomes a universal input $M = N$, the elasticity θ_0 drops out of the formula because producers are uniformly exposed to energy:

$$\frac{d^2 \log Y}{d \log A_e^2} = \left(1 - \frac{1}{\theta_1}\right) \Lambda_e (1 - \Lambda_e).$$

Note that even in this case, the formula is different from that of the case of a single factor where the first term on the right-hand side is $\theta_1 - 1$ instead of $1 - 1/\theta_1$. This reflects the

aforementioned fact that in contrast to the case of a single factor, labor cannot be reallocated to the production of energy following a negative shock to energy, which further amplifies the negative impact of the energy shock.

Macro-Influence—Multiple Factors

In contrast to the case of a single factor, shocks to prices now propagate downstream and upstream. The result derived in Proposition 8 when there is only one factor breaks down when there are multiple factors: two producers with identical demand chains do not necessarily have the same importance. This is because they might have different direct and indirect exposures to the different factors. As a result, the role of a producer as a consumer matters in addition to its role as a supplier.

5. BEYOND CES

The input–output covariance operator defined in equation (5) is a key concept capturing the substitution patterns in economies where all production and utility functions are nested-CES functions. In this section, we generalize this input–output covariance operator in such a way that allows us to work with arbitrary production functions.

For a producer k with cost function \mathbf{C}_k , we define the Allen–Uzawa elasticity of substitution between inputs x and y as

$$\theta_k(x, y) = \frac{\mathbf{C}_k d^2 \mathbf{C}_k / (dp_x dp_y)}{(d\mathbf{C}_k / dp_x)(d\mathbf{C}_k / dp_y)} = \frac{\epsilon_k(x, y)}{\Omega_{ky}},$$

where $\epsilon_k(x, y)$ is the elasticity of the demand by producer k for input x with respect to the price p_y of input y , and Ω_{ky} is the expenditure share in cost of input y .

Note the following properties. Because of the symmetry of partial derivatives, we have $\theta_k(x, y) = \theta_k(y, x)$. Because of the homogeneity of degree 1 of the cost function in the prices of inputs, we have the homogeneity identity $\sum_{1 \leq y \leq N+1+F} \Omega_{ky} \theta_k(x, y) = 0$.

We define the *input–output substitution operator* for producer k as

$$\Phi_k(\Psi_{(i)}, \Psi_{(j)}) = - \sum_{1 \leq x, y \leq N+1+F} \Omega_{kx} [\delta_{xy} + \Omega_{ky} (\theta_k(x, y) - 1)] \Psi_{xi} \Psi_{yj} \quad (11)$$

$$= \frac{1}{2} E_{\Omega^{(k)}} ((\theta_k(x, y) - 1) (\Psi_i(x) - \Psi_i(y)) (\Psi_j(x) - \Psi_j(y))), \quad (12)$$

where δ_{xy} is the Kronecker delta, $\Psi_i(x) = \Psi_{xi}$ and $\Psi_j(x) = \Psi_{xj}$, and the expectation on the second line is over x and y . The second line can be obtained from the first using the symmetry of Allen–Uzawa elasticities of substitution and the homogeneity identity.

In the CES case with elasticity θ_k , all the cross Allen–Uzawa elasticities are identical with $\theta_k(x, y) = \theta_k$ if $x \neq y$, and the own Allen–Uzawa elasticities are given by $\theta_k(x, x) = -\theta_k(1 - \Omega_{kx})/\Omega_{kx}$. It is easy to verify that we then recover the input–output covariance operator:

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

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

Luckily, it turns out that all of the results stated so far can be generalized to non-CES economies simply by replacing terms of the form $(\theta_k - 1)\text{Cov}_{\Omega(k)}(\Psi_i, \Psi_j)$ by $\Phi_k(\Psi_i, \Psi_j)$. For example, equation (8) in Proposition 9 becomes

$$\frac{d^2 \log Y}{d \log A_j d \log A_i} = \sum_{k=0}^N \Phi_k(\Psi_i, \Psi_j) - \sum_{f=N+1}^{N+F} \frac{d \log A_f}{d \log A_j} \sum_{k=0}^N \Phi_k(\Psi_i, \Psi_f).$$

By replacing the input–output covariance operator with the input–output substitution operator, we fully characterize the Hessian of the output function of the general economy described in Section 2, with arbitrary, and potentially, non-homothetic production functions, an arbitrary number of factors, and arbitrary patterns of input–output linkages.

Intuitively, $\Phi_k(\Psi_i, \Psi_j)$ captures the way in which k redirects demand expenditure towards i in response to proportional unit decline in the price of j . To see this, we make use of the following observation: the elasticity of the expenditure share of producer k on input x with respect to the price of input y is given by $\delta_{xy} + \Omega_{ky}(\theta_k(x, y) - 1)$. Equation (11) requires considering, for each pair of inputs x and y , how much the proportional reduction Ψ_{yj} in the price of y induced by a unit proportional reduction in the price of j causes producer k to increase its expenditure share on x (as measured by $-\Omega_{kx}[\delta_{xy} + \Omega_{ky}(\theta_k(x, y) - 1)]\Psi_{yj}$) and how much x is exposed to i (as measured by Ψ_{xi}).

Equation (12) says that this amounts to considering, for each pair of inputs x and y , whether or not increased exposure to j as measured by $\Psi_j(x) - \Psi_j(y)$, corresponds to increased exposure to i as measured by $\Psi_i(x) - \Psi_i(y)$, and whether x and y are complements or substitutes as measured by $(\theta_k(x, y) - 1)$. If x and y are substitutes, and $\Psi_j(x) - \Psi_j(y)$ and $\Psi_i(x) - \Psi_i(y)$ are both positive, then substitution across x and y by k , in response to a shock to a decrease in the price of j , increases demand for i .

6. QUANTITATIVE ILLUSTRATION

In this section, we develop some illustrative quantitative applications of our results to gauge the practical importance of the nonlinearities that we have identified. We perform three exercises focusing on macroeconomic phenomena at different frequencies. First, we calibrate a multi-sector business-cycle model with sectoral productivity shocks. We match the observed input–output data, and use the best available information to choose the structural (micro) elasticities of substitution, and we match the volatility of sectoral shocks at business-cycle frequencies. We compare the outcome of the nonlinear model to its first-order approximation. In the second exercise, we study the macroeconomic impact of the energy crisis of the 1970s using a nonparametric generalization of [Hulten \(1978\)](#) that takes second-order terms into account. In the third and final exercise, we investigate the importance of nonlinearities which underpin Baumol’s cost disease for long-run aggregate TFP growth. All our exercises suggest that production is highly nonlinear.

6.1. A Quantitative Multi-Sector Business-Cycle Model With Sectoral Productivity Shocks

In this section, we use a simple version of the structural model defined in Section 2, calibrate it with sectoral shocks at business-cycle frequencies, and solve for the equilibrium

using global solution methods.³⁸ The final demand function is

$$\frac{Y}{\bar{Y}} = \left(\sum_{i=1}^N \omega_{0i} \left(\frac{c_i}{\bar{c}_i} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}.$$

The production function of industry i is

$$\frac{y_i}{\bar{y}_i} = A_i \left(\omega_{il} \left(\frac{l_i}{\bar{l}_i} \right)^{\frac{\theta-1}{\theta}} + (1 - \omega_{il}) \left(\frac{\hat{X}_i}{\bar{X}_i} \right)^{\frac{\theta-1}{\theta}} \right)^{\frac{\theta}{\theta-1}},$$

consisting of labor inputs l_i and intermediate inputs \hat{X}_i .

We consider two polar opposite possibilities for the labor market: the case where each labor type is specific to each industry and cannot be reallocated, and the case where there is a common factor which can be reallocated across all industries. In light of increasing evidence (see, e.g., Acemoglu, Autor, Dorn, Hanson, and Price (2016), Autor, Dorn, and Hanson (2016), Notowidigdo (2011)) that labor is not easily reallocated across industries or regions after shocks in the short run, we view the no-reallocation case as more realistic for modeling the short-run impact of shocks, and the full-reallocation case as better suited to study the medium- to long-run impact shocks.³⁹

The composite intermediate input X_i is given by

$$\frac{X_i}{\bar{X}_i} = \left(\sum_{j=1}^N \omega_{ij} \left(\frac{x_{ij}}{\bar{x}_{ij}} \right)^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon-1}},$$

where x_{ij} are intermediate inputs from industry j used by industry i .

Data and Calibration

We work with the 88 sector U.S. KLEMS annual input–output data from Dale Jorgenson and his collaborators, dropping the government sectors. The data set contains sectoral output and inputs from 1960 to 2005. We use the sector-level TFP series computed by Carvalho and Gabaix (2013) using the methodology of Jorgenson, Gollop, and Fraumeni (1987).

³⁸For comparison, Table III in Appendix D of the Supplemental Material shows the macro-moments for the model using the second-order approximation. The second-order approximation does a very good job at capturing the mean and standard deviation of aggregate output at both annual and quadrennial frequencies. It also performs well for skewness and kurtosis at an annual frequency, but less so at a quadrennial frequency. Basically, the quality of the approximation is worse for larger shocks and higher moments of output. Of course, whenever it is feasible to solve the model nonlinearly, the fully nonlinear solution is preferable. However, even in these cases, our analytical pen-and-paper approach also sheds light on the mechanisms driving nonlinearities that would be lost were we to simply solve a series of nonlinear equations on a computer. Furthermore, our sufficient statistics approach allows us to construct the second-order approximation without needing to fully specify the nonlinear model.

³⁹In our baseline calibrations, we assume that intermediate inputs can be freely reallocated across producers even in the short run. This is sensible since intermediate goods are probably easier to reallocate than labor. We refer the reader to Appendix C of the Supplemental Material for a version of these calibrations in the presence of costs of adjusting intermediate inputs. These adjustment costs hamper the reallocation of intermediate inputs across producers and magnify the nonlinearities of the model.

We calibrate the expenditure share parameters to match the input–output table, using 1982 (the middle of the sample) as the base year. We specify sectoral TFP shocks to be log-normally distributed so that $\log A_i \sim \mathcal{N}(-\Sigma_{ii}/2, \Sigma_{ii})$, where Σ_{ii} is the sample variance of log TFP growth for industry i . We work with uncorrelated sectoral shocks since the average correlation between sectoral growth rates is small (less than 5%). Our results are not significantly affected if we matched the whole covariance matrix of sectoral TFP instead.

We consider shocks at annual and quadrennial horizons, the latter corresponding to the average period of a business cycle. The average standard deviation at a quadrennial frequency is about twice its value at an annual frequency. In fact, it is the only difference between the annual and quadrennial calibrations. Nonlinearities, which matter more for bigger shocks, are more important at a quadrennial frequency than at an annual frequency.

Our specification assumes only three distinct structural microeconomic elasticities of substitution. This is because estimates of more disaggregated elasticities are not available. For our benchmark calibration, we set $(\sigma, \theta, \varepsilon) = (0.9, 0.5, 0.001)$. We set the elasticity of substitution in consumption $\sigma = 0.9$, following [Atalay \(2017\)](#), [Herrendorf, Rogerson, and Valentinyi \(2013\)](#), and [Oberfield and Raval \(2014\)](#), all of whom used an elasticity of substitution in consumption (across industries) of slightly less than 1. For the elasticity of substitution across value-added and intermediate inputs, we set $\theta = 0.5$. This accords with the estimates of [Atalay \(2017\)](#), who estimated this parameter to be between 0.4 and 0.8, as well as [Boehm, Flaaen, and Pandalai-Nayar \(2017\)](#), who estimated this elasticity to be close to zero. Finally, we set the elasticity of substitution across intermediate inputs to be $\varepsilon = 0.001$, which matches the estimates of [Atalay \(2017\)](#).

Owing to uncertainty surrounding the estimates for $(\sigma, \theta, \varepsilon)$, we include many robustness checks in Tables IV–VII in Appendix D of the Supplemental Material. In the main text in Table I, we focus on four sets of elasticities $(\sigma, \theta, \varepsilon)$: our benchmark calibration (0.9, 0.5, 0.001); a calibration with lower but still plausible elasticities (0.7, 0.3, 0.001); a calibration with higher elasticities (0.9, 0.6, 0.2); and a (close to) Cobb–Douglas calibration (0.99, 0.99, 0.99). In Appendix D of the Supplemental Material, we report robustness checks for different values of these elasticities on a grid with $\sigma \in \{0.8, 0.9, 0.99\}$, $\theta \in \{0.4, 0.5, 0.6, 0.99\}$, and $\varepsilon \in \{0.001, 0.2, 0.99\}$. Our results in Table I are not sensitive to the exact value of $(\sigma, \theta, \varepsilon)$ provided that the elasticities are collectively low enough that the calibration matches the observed volatility of the Domar weights.

Since the volatility of Domar weights is a measure of the size of the second-order terms in the model, we use the volatility of Domar weights as a sanity check for our calibration. Specifically, we target $\sigma_\lambda = \sum_i \bar{\lambda}_i \sigma_{\lambda_i}$, where $\bar{\lambda}_i$ is the time-series average of the i th Domar weight and σ_{λ_i} is the time-series standard deviation of industry i 's log-differenced Domar weight. In our data, at annual frequency, $\sigma_\lambda \approx 0.13$, and for quadrennial frequency, $\sigma_\lambda \approx 0.27$. With no labor reallocation, our baseline calibrations match these numbers relatively well; the lower-elasticity calibrations overshoot; the higher-elasticity calibrations undershoot; and the Cobb–Douglas calibrations deliver zero volatility of the Domar weights. Allowing for labor reallocation reduces the volatility of the Domar weights.⁴⁰

⁴⁰To match the data, we assume that changes in sectoral TFPs are the sole driver of fluctuations in the Domar weight. Other types of shocks, besides industry-level TFP shocks, could also drive volatility to the Domar weights (e.g., shocks to the composition of demand). While such shocks, whatever they are, would also indicate the presence of nonlinearities, they would imply different elasticities of substitution for the calibration of the micro-elasticities of substitution. We abstract from this issue in calibrating the parameters of the model in Section 6.1.

Results

Table I displays the mean, standard deviation, skewness, and excess kurtosis of log aggregate output for various specifications. For comparison, the table also shows these moments for aggregate TFP growth in the data.⁴¹ In addition, we also report the volatility of the Domar weights, both in the model and in the data. Since we have too few annual (and even fewer quadrennial) observations of aggregate TFP, we do not report the skewness and excess kurtosis: the implied confidence intervals for these estimates would be so large as to make the point estimates uninformative. In Appendix D of the Supplemental Material, we report results for numerous other permutations of the elasticities of substitution for robustness.

Overall, given our elasticities of substitution, the model with full reallocation is unable to replicate the volatility of the Domar weights at either annual or quadrennial frequency, suggesting that this model is not nonlinear enough to match the movements in the Domar weights as arising from sectoral productivity shocks. On the other hand, the model without reallocation is able to match the volatility of the Domar weights, which is consistent with the intuition of Section 3.

TABLE I
SIMULATED AND ESTIMATED MOMENTS^a

$(\sigma, \theta, \epsilon)$	Mean	Std	Skew	Ex-Kurtosis	σ_λ
Full Reallocation—Annual					
(0.7, 0.3, 0.001)	−0.0023	0.011	−0.10	0.1	0.090
(0.9, 0.5, 0.001)	−0.0022	0.011	−0.08	0.0	0.069
(0.9, 0.6, 0.2)	−0.0020	0.011	−0.05	0.0	0.056
(0.99, 0.99, 0.99)	−0.0013	0.011	0.01	0.0	0.001
No Reallocation—Annual					
(0.7, 0.3, 0.001)	−0.0045	0.012	−0.31	0.4	0.171
(0.9, 0.5, 0.001)	−0.0034	0.012	−0.18	0.1	0.115
(0.9, 0.6, 0.2)	−0.0024	0.011	−0.11	0.1	0.068
(0.99, 0.99, 0.99)	−0.0011	0.011	0.00	0.0	0.001
Full Reallocation—Quadrennial					
(0.7, 0.3, 0.001)	−0.0118	0.026	−0.4	0.4	0.307
(0.9, 0.5, 0.001)	−0.0113	0.026	−0.28	0.4	0.176
(0.9, 0.6, 0.2)	−0.0100	0.026	−0.23	0.2	0.133
(0.99, 0.99, 0.99)	−0.0058	0.025	0.01	0.0	0.003
No Reallocation—Quadrennial					
(0.7, 0.3, 0.001)	−0.0270	0.037	−2.18	12.7	0.404
(0.9, 0.5, 0.001)	−0.0187	0.030	−1.11	3.6	0.267
(0.9, 0.6, 0.2)	−0.0129	0.027	−0.44	0.7	0.154
(0.99, 0.99, 0.99)	−0.0057	0.025	0.00	0.0	0.002
Annual Data	—	0.015	—	—	0.13
Quadrennial Data	—	0.030	—	—	0.27

^aFor the data, we use the demeaned growth rate of aggregate TFP. For the model, we use the sample moments of log output. The simulated moments are calculated from 50,000 draws. Skewness and kurtosis of the data are impossible to estimate with enough precision and so we do not report them.

⁴¹Since our model has inelastic factor supply, its output is more comparable to aggregate TFP than to real GDP. As shown by Gabaix (2011) and Carvalho and Gabaix (2013), elastic capital and labor supply would further amplify TFP shocks.

Let us consider each moment in turn, starting with the mean. For our benchmark calibration $(\sigma, \theta, \varepsilon) = (0.9, 0.5, 0.001)$, the model without reallocation matches the volatility of Domar weights at both annual and quadrennial frequency. The reductions in the mean are around 0.3% at annual and just under 2% at quadrennial frequency. Since our log-normal shocks are calibrated to have a mean of 1 in levels, as we increase the variance, the mean of the log declines. We can net out this mechanical effect by subtracting the average loss in performance from the ones in the log-linear Cobb–Douglas model. At an annual frequency, this results in a loss from nonlinearities of $0.34\% - 0.11\% = 0.23\%$, and at a quadrennial frequency, the loss from nonlinearities is $1.87\% - 0.57\% = 1.30\%$. These numbers, which identify the welfare costs of sectoral shocks arising from concavity in production, can be compared with the welfare costs of fluctuations à la Lucas (1987) arising from concavity in utility, which are around 0.01% and 0.05%, respectively, in these calibrations. The reduction in the mean increases as we pump up the degree of nonlinearity, and so do the corresponding losses from nonlinearities.

A qualitatively similar pattern holds for the model with full reallocation. In that case, the reductions in the mean are smaller: $0.2\% - 0.1\% = 0.1\%$ at an annual frequency and $1.1\% - 0.57\% = 0.46\%$ at a quadrennial frequency. Unsurprisingly, the approximately Cobb–Douglas model behaves similarly regardless of the mobility of labor—this follows from the fact that Hulten’s theorem holds globally for a Cobb–Douglas model where no labor reallocation takes place in equilibrium whether or not it is allowed.

At an annual frequency, both models undershoot a bit on the overall volatility of aggregate TFP. One reason why the model undershoots on standard deviation, particularly at annual frequency, is that we restrict the industry-level productivity shocks to be independent, whereas in the data, particularly at higher frequencies, they feature some correlation. At a quadrennial frequency, the model better matches the volatility of aggregate TFP. The degree of nonlinearity makes little difference for the volatility of aggregate TFP at an annual frequency.⁴² Nonlinearities matter more for the volatility of aggregate TFP at a quadrennial frequency because the shocks are larger.

Finally, the models with and without reallocation both generate negative skewness and some positive excessive kurtosis (in fact, a very high amount for the quadrennial specifications without reallocation). The skewness and excess kurtosis fatten the left tail of the distribution, providing an endogenous explanation for “rare disasters.” Unlike Acemoglu, Ozdaglar, and Tahbaz-Salehi (2017) or Barro (2006), to achieve rare disasters, we do not need to assume fat-tailed exogenous shocks nor rule out “rare bonanzas” a priori. Instead, these features are endogenously generated by the nonlinearities in the model. This can be seen in Figure 5, where we plot the histograms for the benchmark calibrations with no reallocation (which match the volatility of Domar weights) and for a log-linear approximation subject to the same shocks. As expected, nonlinearities are more important at a quadrennial frequency than at an annual frequency because the shocks are larger (more volatile).

We also consider the response of aggregate output to shocks to specific industries, using our benchmark calibration. It turns out that for a large negative shock, the “oil and gas” industry produces the largest negative response in aggregate output, despite the fact it is

⁴²Whereas for the mean, skewness, and kurtosis, the second-order terms are the dominant power in the Taylor expansion (since the linear terms have no effect), for the variance, the dominant power is the linear term. For example, letting σ be the standard deviation of the shocks, the approximation of the variance in Appendix D of the Supplemental Material shows that the contribution of the linear terms scales in σ^2 , whereas that of the nonlinear terms scales in σ^4 .

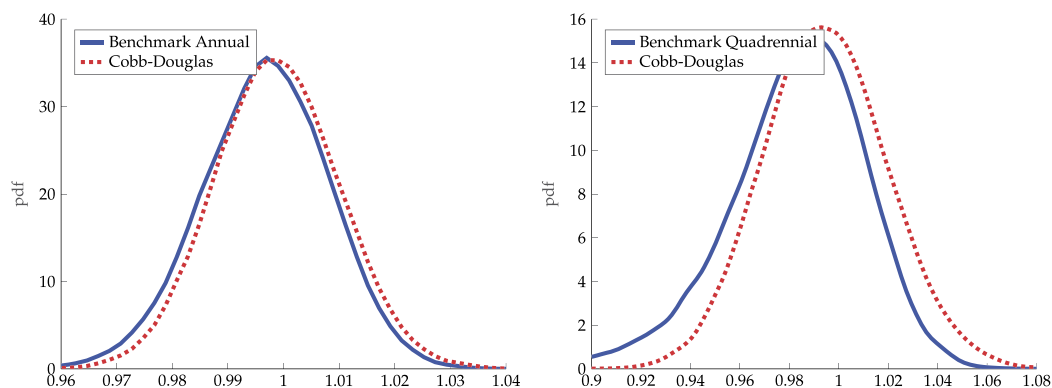


FIGURE 5.—The left panel shows the distribution of aggregate output for the benchmark model and log-linear model at an annual frequency. The right panel shows these for shocks at a quadrennial frequency. The difference between the two frequencies is that shocks are larger (their volatility is higher) at a quadrennial than at an annual frequency. The benchmark model has $(\sigma, \theta, \varepsilon) = (0.9, 0.5, 0.001)$ and no labor reallocation. Note that the scales are different in these two figures.

not the largest industry in the economy. Figure 6 plots the response of aggregate output for shocks to the oil and gas industry as well as for the “retail trade (excluding automobiles)” industry. The retail trade industry has a similar sales share, and therefore, to a first order, both industries are equally important. As expected, the nonlinear model amplifies negative shocks and mitigates positive shocks. However, whereas output is roughly log-linear for shocks to retail trade, output is highly nonlinear with respect to shocks to oil and gas.

The intuition for this asymmetry comes from the examples in Figures 3 and 4. “Oil and gas” is an approximately universal input, so that the downstream elasticity of substitution (in consumption) σ is less relevant and the upstream elasticity of substitution (in produc-

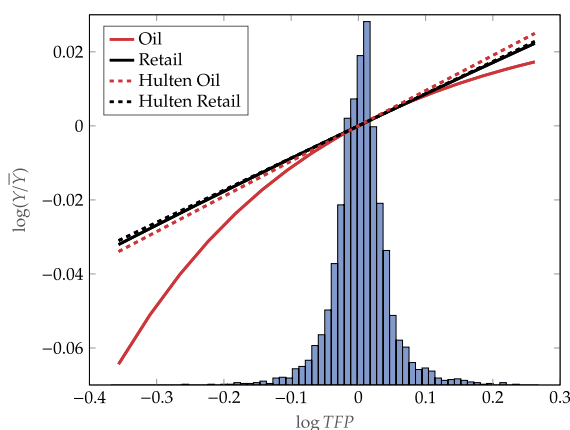


FIGURE 6.—The effect of TFP shocks to the oil and gas industry and the retail trade industry. Both industries have roughly the same sales share, and so they are equally important up to a first-order approximation (dotted line). The nonlinear model is more fragile to both shocks than the log-linear approximation. The “oil and gas” industry is significantly more important than “retail trade” for large negative shocks. The histogram is the empirical distribution of sectoral annual TFP shocks pooled over the whole sample. The model has $(\sigma, \theta, \varepsilon) = (0.9, 0.5, 0.001)$ with no labor reallocation and no adjustment costs.

tion) θ is more relevant. Since $\theta \ll \sigma$ in our calibration, this means that output is more nonlinear in shocks to “oil and gas” than in shocks to “retail trade.” Furthermore, “oil and gas” have a relatively low share of intermediate input usage. As a result, in response to a negative shock, resources cannot be reallocated to boost the production of oil and gas. This means that we are closer to the economy depicted in Figure 4 than the one in Figure 3, with a correspondingly lower GE elasticity of substitution between oil and gas and everything else due to the lack of reallocation.

The strong asymmetry between the effects of positive and negative shocks is consistent with the empirical findings of Hamilton (2003) that oil price increases are much more important than oil price decreases.⁴³

6.2. The Effect of Oil Shocks

In this section, we use the oil shocks of the 1970s to demonstrate the way nonlinearities can amplify the macroeconomic impact of industry-level shocks.⁴⁴ To recap the history, starting in 1973, coordinated action by OPEC caused the price of crude oil to increase from \$3.5 per barrel in 1972 to \$11 per barrel in 1974. In 1979, OPEC implemented a second round of quantity restrictions which caused the price of crude to soar to \$31 per barrel. Shortly after this, the Iranian revolution of 1979 and the ensuing Iraqi invasion of Iran caused further disruptions to global crude oil supply. The price peaked at \$37 in 1980. Starting in the early 1980s, with the abdication of the Shah, OPEC’s pricing structure collapsed and, in a bid to maintain its market share, Saudi Arabia flooded the market with inexpensive oil. In real terms, the price of crude oil declined back to its pre-crisis levels by 1986.

We adopt a nonparametric approach which allows us to account for the second-order macroeconomic impact of microeconomic shocks using ex post data. Instead of trying to predict how Domar weights change in response to a shock as would be required for a counterfactual exercise, we can simply observe it. Formally, we rely on the following result.

PROPOSITION 10: *Up to the second order in the vector Δ , we have*

$$\log(Y(A + \Delta)/Y(A)) = \frac{1}{2}[\lambda(A + \Delta) + \lambda(A)]'(\log(A + \Delta) - \log(A)).$$

⁴³Figure 6 may give the impression that the relative ranking of industries is stable as a function of the size of the shock. The oil industry is always more important than the retail trade industry for negative shocks, and always less important for positive shocks. However, this need not be the case. In Appendix D of the Supplemental Material, we plot aggregate output as a function of shocks to the “oil and gas” industry and the construction industry. The construction industry is larger than the oil industry. Therefore, the first-order approximation implies that it should be more important. The nonlinear model also behaves the same way for positive shocks, and small negative shocks. However, for very large negative shocks, the “oil and gas” industry once again becomes more important.

⁴⁴Although our structural model suggests that the “oil and gas” extraction industry is important, it abstracts away from trade, by assuming all intermediate inputs are sourced domestically, with net imports showing up only in final demand. Hence, the Domar weight of the “oil and gas” industry measures domestic production, rather than domestic consumption. Since the oil price shocks did not directly affect the productivity of domestic oil production, this means that they are not measured in our sectoral TFP data (which is for domestic production). Furthermore, our industry classification is too coarse to isolate crude oil separately from other petrochemicals. For this reason, we use global (rather than U.S.) data.

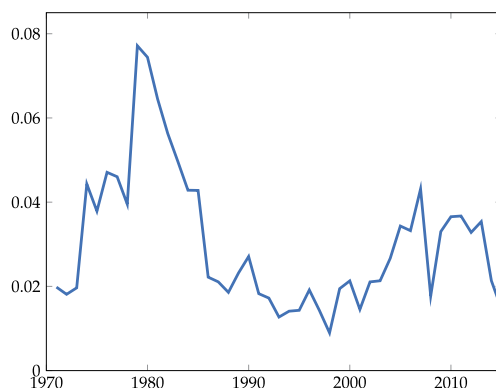


FIGURE 7.—Global expenditures on crude oil as a fraction of world GDP.

The idea of averaging weights across two periods is due to [Törnqvist \(1936\)](#). Proposition 10 relates the macroeconomic impact of microeconomic shocks to the size of the shock and the corresponding Domar weights before and after the shock.⁴⁵

We measure the price of oil using the West Texas Intermediate Spot Crude Oil price from the Federal Reserve Database. Global crude oil production, measured in thousand tonne of oil equivalents, is from the OECD. World GDP, in current USD, is from the World Bank national accounts data. The choice of the pre and post Domar weight is not especially controversial. Crude oil, as a fraction of world GDP, increased from 1.8% in 1972 to 7.6% in 1979. Reassuringly, the Domar weight is back down to its pre-crisis level by 1986 (see Figure 7). This means that, taking the second-order terms into account, we need to weight the shock to the oil industry by $1/2(1.8\% + 7.6\%) = 4.7\%$. Hence, the second-order terms amplify the shock by a factor of $4.7/1.8 \approx 2.6$.

Calibrating the size of the shock to the oil industry is more tricky, since it is not directly observed. If we assume that oil is an endowment, then we can simply measure the shock via changes in the physical quantity of production. To do this, we demean the log growth rate in global crude oil production, and take the shock to be the cumulative change in demeaned growth rates from 1973 to 1980, which gives us a shock of -13% .⁴⁶

⁴⁵One can always compute the full nonlinear impact of a shock on output by computing $\int_{\tilde{A}}^{\tilde{A}+\Delta} \lambda(\tilde{A}) d \log \tilde{A}$, and our formula approximates this integral by performing a first-order (log) approximation of the Domar weight $\lambda(\tilde{A})$, or equivalently, a second-order (translog) approximation of aggregate output. In theory, if TFP is a continuous diffusion, then one can disaggregate time periods and compute the impact of shocks over a time period $[t, t + \delta]$ as $\int_t^{t+\delta} \lambda(A_s) d \log(A_s)$ which can be seen as a repeated application of Hulten's theorem at every point in time over infinitesimal intervals of time. However, when TFP has jumps, then this decomposition no longer applies. In any case, even when it does apply, and when the required high-frequency data regarding TFP shocks and Domar weights are available, it can only be useful ex post to assess the changes in aggregate output over an elapsed period of time due to the TFP shocks $d \log(A_s)$ to a given sector given the observed path of Domar weight $\lambda(A_s)$. It is of no use ex ante to predict how these future shocks will affect aggregate output because one would need to know how the Domar weight will change over time as a result of the shocks, and hence of no use to run counterfactuals. This latter part is precisely what the second-order approximation at the heart of our paper accomplishes.

⁴⁶We use the demeaned growth rate to remove the overall (positive) trend in production. Intuitively, if everything is growing at the same rate, then a negative oil shock is a reduction in the growth of oil relative to trend. Of course, one can easily quibble with this estimate of the size of the shock, but fortunately, the degree of amplification (defined as the ratio of the second-order approximation to the first) is independent of

Putting this all together, the first-order impact on aggregate output is therefore

$$1.8\% \times -13\% = -0.23\%$$

On the other hand, the second-order impact on aggregate output is

$$\frac{1}{2}(1.8\% + 7.6\%) \times -13\% = -0.61\%.$$

Hence, accounting for the second-order terms amplifies the impact of the oil shocks significantly, so that oil shocks can be macroeconomically significant even without any financial or demand side frictions.⁴⁷

6.3. *Baumol's Cost Disease and Long-Run Growth*

Our final empirical exercise looks to quantify the importance of nonlinearities on long-run productivity growth. For this exercise, we use World KLEMS data for the United States from 1948 to 2014.

The “nonlinear” measure of aggregate TFP growth over the sample is built by updating the Domar weights every period:

$$\Delta \log \text{TFP}^{\text{nonlinear}} = \sum_{i=1}^N \sum_{t=1948}^{2013} \lambda_{i,t} (\log A_{i,t+1} - \log(A_{i,t})).$$

This provides an approximation, by discrete left Riemann sums, of the exact aggregate TFP growth, given by the sum of continuous integrals $\sum_{i=1}^N \int_{1948}^{2014} \lambda_{i,t} d \log A_{i,t}$.

Now consider the following counterfactual: imagine that the economy is log-linear so that the Domar weights are constant throughout the sample at their 1948 values. Assuming that the path for industry-level TFP is unchanged, aggregate TFP growth over the sample would be given by

$$\Delta \log \text{TFP}^{\text{1st order}} = \sum_{i=1}^N \lambda_{i,1948} (\log A_{i,2014} - \log A_{i,1948}).$$

By comparing actual aggregate TFP growth and TFP growth in the counterfactual log-linear economy, we quantify the importance of Baumol's cost disease (Baumol (1967)): the notion that over time, because of complementarities, the sales shares of low-productivity-growth industries increase while those of high-productivity-growth industries decrease, thereby slowing down aggregate TFP growth.

By the end of the sample, aggregate TFP growth in the counterfactual log-linear or Cobb–Douglas economy is around 87% whereas actual TFP grew by 68%. Hence the

our estimate for the size of the shock. So, for any value of the shock, the second-order approximation almost triples the impact of the shock.

⁴⁷As noted by Hamilton (2013), first-order approximations of efficient models assign a relatively small impact to oil price shocks. Hence, the literature has tended to focus on various frictions that may account for the strong statistical relationship between oil shocks and aggregate output. Our calculation suggests that nonlinearities in production may help explain the outsized effect of oil shocks even in efficient models. Furthermore, our calculation makes no allowance for amplification of shocks through endogenous labor supply and capital accumulation, which are the standard channels for amplification of shocks in the business-cycle literature. Hence, coupled with the standard amplification mechanisms of those models, we would expect the reduction in aggregate output to be even larger.

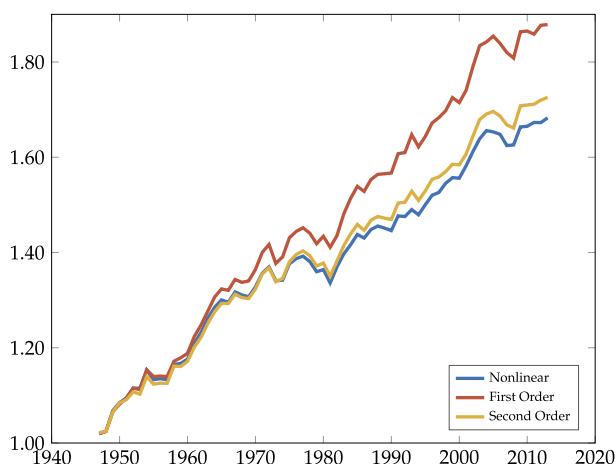


FIGURE 8.—Cumulative change in TFP: nonlinear (actual), first-order approximation, and second-order approximation.

presence of nonlinearities slowed down aggregate TFP growth by around 19 percentage points over the sample period.

Baumol's cost disease is a manifestation of nonlinearities, since $\Delta \log \text{TFP}^{\text{1st order}}$ is a first-order approximation of actual aggregate TFP growth. A second-order approximation, which captures some of the nonlinearities, is given by

$$\Delta \log \text{TFP}^{\text{2nd order}} = \frac{1}{2} \sum_{i=1}^N (\lambda_{i,1948} + \lambda_{i,2014}) (\log A_{i,2014} - \log A_{i,1948}).$$

In Figure 8, we plot aggregate TFP growth using the nonlinear, first-order, and second-order measures. The gap between the nonlinear measure and the first-order approximation is sizeable, but that between the nonlinear measure and the second-order approximation is much smaller.

These findings are in some sense more general than Baumol's hypothesized mechanism. Traditionally, the mechanism for Baumol's cost disease is through prices: because of complementarities, as the prices of high-productivity-growth industries fall, their shares in aggregate output fall, resources are reallocated away from them, thereby reducing aggregate TFP growth. Alternative stories for structural transformation emphasize other mechanisms, principally, non-homotheticities whereby as real income increases, consumers increase their expenditures on industries that happen to have lower productivity growth.⁴⁸ From the perspective of our framework, non-homotheticities and complementarities are both nonlinearities. Our calculations of the impact of nonlinearities encompass both types of mechanisms.^{49,50}

⁴⁸See Herrendorf, Rogerson, and Valentinyi (2013) as an example.

⁴⁹Our theoretical characterizations cover all these nonlinearities. In fact, formally, non-homotheticities can always be represented via non-unitary elasticities of substitution between inputs and a fixed factor. To see this, note that any non-homothetic function $f(x)$ can be extended into a constant-returns function $\tilde{f}(x, y) = yf(x/y)$ where $f(x) = \tilde{f}(x, 1)$. Then, non-homotheticity in f is equivalent to a non-unitary elasticity of substitution between x and y in $\tilde{f}(x, y)$.

⁵⁰Another potential source of nonlinearities is shocks to the composition of demand. For example, a Cobb–Douglas model with shocks to the share parameters is a nonlinear model since the cross-partial derivatives

7. CONCLUSION

The paper points to many unanswered questions. For instance, it shows that the macroeconomic impact of a microeconomic shock depends greatly on how quickly factors can be reallocated across production units. Since our structural model is static, we are forced to proxy for the temporal dimension of reallocation by resorting to successive comparative statics. In ongoing work, we investigate the dynamic adjustment process more rigorously and find that although we can think of the no-reallocation and perfect-reallocation cases as the beginning and end of the adjustment, the speed of adjustment also greatly depends on the microeconomic details. This means that the dynamic response of aggregate output to different shocks is greatly affected by issues like geographic or sectoral mobility of labor, even with perfect and complete markets that allow us to abstract from distributional issues. Our structural application also lacks capital accumulation and endogenous labor supply, and incorporating these into the present analysis is an interesting area for future work.

Furthermore, in this paper, we have focused exclusively on the way shocks affect aggregate output. In [Baqaei and Farhi \(2018a\)](#), we focused on how shocks affect factor shares and derived characterizations of the macroeconomic elasticities of substitution between factors and of the macroeconomic bias of technical change. In [Baqaei and Farhi \(2018b\)](#), we took up the related question of how shocks affect non-aggregate outcomes, namely, how shocks propagate from one producer to another, and how microeconomic variables comove with one another in a production network.

In addition, this paper assumes that aggregate final demand is homothetic. Non-homothetic aggregate final demand can arise from non-homothetic individual final demand or from the aggregation of heterogeneous but homothetic individual demands. In [Baqaei and Farhi \(2018b, 2019\)](#), we showed how to take these elements into account. In particular, in [Baqaei and Farhi \(2018b\)](#), we show how to combine input–output production networks with heterogeneous agents and non-homothetic final demand in closed economies. In [Baqaei and Farhi \(2019\)](#), we showed how to take these elements into account in trade models.

Finally, this paper assumes away non-technological frictions, but the forces we identify do not disappear in richer models with inefficient equilibria. Non-unitary elasticities of substitution in networks can amplify or attenuate the underlying distortions. In [Baqaei and Farhi \(2018c\)](#), we undertook a systematic characterization of these effects. We showed that in inefficient models, the “second-order” terms that we characterize in this paper can become first order.

APPENDIX A: PROOFS

PROOF OF THEOREM 1: Since the first welfare theorem holds, the equilibrium allocation solves

$$Y(A_1, \dots, A_N) = \max_{c_i, x_{ij}, l_{ij}} \mathcal{D}(c_1, \dots, c_n) + \sum_i \mu_i \left(A_i F_i((l_{ij})_j, (x_{ij})_j) - \sum_j x_{ij} - c_i \right) \\ + \sum_i \lambda_i \left(\bar{l}_i - \sum_j l_{ji} \right),$$

of aggregate output with respect to industry TFP and share shocks are nonzero. Our results also cover such economies.

where \bar{l}_i is the endowment of each labor type, and μ_i and λ_i are Lagrange multipliers. The envelope theorem then implies that

$$\frac{dY}{dA_i} = \mu_i F_i((l_{ij})_j, (x_{ij})_j) = \mu_i y_i.$$

If we show that μ_i is equal to the price of i in the competitive equilibrium, then we are done.

Meanwhile, for each good j , either there exists another producer i using that good as an input, or the household must consume that input (otherwise, the input is irrelevant and has a price of zero). Hence, in a competitive equilibrium, we must have either

$$p_i \frac{\partial F_i}{\partial x_{ij}} = p_j \quad (13)$$

and/or

$$P_c \frac{\partial \mathcal{D}}{\partial c_j} = p_j, \quad (14)$$

where P_c is the ideal price index associated with \mathcal{D} (which we can take to be the numeraire). The expression above uses the fact that \mathcal{D} is constant-returns-to-scale.

On the other hand, the first-order conditions of the social planners problem implies that for each j , either

$$\mu_i \frac{\partial F_i}{\partial x_{ij}} = \mu_j \quad (15)$$

and/or

$$\frac{\partial \mathcal{D}}{\partial c_j} = \mu_j. \quad (16)$$

Hence $\mu_i = p_i$ for every i .

Q.E.D.

PROOF OF THEOREM 2: Differentiate $\sum_i \lambda_i = \xi$ to get

$$\begin{aligned} \lambda_i \frac{d \log \lambda_i}{d \log A_i} &= \xi \frac{d \log \xi}{d \log A_i} - \sum_{j \neq i} \lambda_j \frac{d \log \lambda_j}{d \log A_i} \\ &= \xi \frac{d \log \xi}{d \log A_i} - \sum_{j \neq i} \lambda_j \frac{d \log \lambda_j / \lambda_i}{d \log A_i} - \sum_{j \neq i} \lambda_j \frac{d \log \lambda_i}{d \log A_i}, \end{aligned}$$

which we can rewrite as

$$\lambda_i \frac{d \log \lambda_i}{d \log A_i} = \xi \frac{d \log \xi}{d \log A_i} + \sum_{j \neq i} \lambda_j \left(1 - \frac{1}{\rho_{ji}} \right) - (\xi - \lambda_i) \frac{d \log \lambda_i}{d \log A_i}.$$

Rearrange this to get

$$\xi \frac{d \log \lambda_i}{d \log A_i} = \xi \frac{d \log \xi}{d \log A_i} + \sum_{j \neq i} \lambda_j \left(1 - \frac{1}{\rho_{ji}} \right). \quad (17)$$

Finally, Theorem 1 implies that

$$\frac{d^2 \log Y}{d \log(A_i)^2} = \lambda_i \frac{d \log \lambda_i}{d \log A_i}.$$

Substitute (17) into the expression above to get the desired result. Last, if Y is homogeneous, Euler's theorem implies that

$$\sum_i \frac{dY}{dA_i} \frac{A_i}{Y} = \sum_i \lambda_i = \xi,$$

hence, $d \log \xi / d \log A_i = 0$.

Q.E.D.

PROOF OF PROPOSITION 4: We prove a slightly more general formulation with arbitrary variance covariance matrix and an arbitrary twice-differentiable utility function:

$$\begin{aligned} E(u(Y(A))) &\approx E\left(u(Y(\bar{A})) + u'(Y(\bar{A}))\nabla Y(\bar{A})(A - \bar{A})\right. \\ &\quad + \frac{1}{2}u''(Y(\bar{A}))(A - \bar{A})'(\nabla Y(\bar{A}) \circ \nabla Y(\bar{A}'))(A - \bar{A}) \\ &\quad \left. + \frac{1}{2}u'(Y(\bar{A}))(A - \bar{A})'\nabla^2 Y(\bar{A})(A - \bar{A})\right) \\ &= u(Y(\bar{A})) + \frac{1}{2}u''(Y(\bar{A}))\text{tr}((\nabla Y(\bar{A}) \circ \nabla Y(\bar{A}'))\Sigma) \\ &\quad + \frac{1}{2}u'(Y(\bar{A}))\text{tr}(\nabla^2 Y(\bar{A})\Sigma). \end{aligned}$$

Now apply Hulten's theorem to get

$$= u(Y(\bar{A})) + \frac{1}{2}u''(Y(\bar{A})) \sum_{k,j} \lambda_k \lambda_j \sigma_{jk} + \frac{1}{2}u'(Y(\bar{A})) \sum_{j,k} \frac{d^2 Y}{dA_k dA_j} \sigma_{jk},$$

with idiosyncratic shocks, this simplifies to

$$= u(Y(\bar{A})) + \frac{1}{2}u''(Y(\bar{A})) \sum_k \lambda_k^2 \sigma_k^2 + \frac{1}{2}u'(Y(\bar{A})) \sum_k \frac{d^2 Y}{dA_k^2} \sigma_k^2.$$

The second summand is the Lucas term (which equals zero when u is linear), and the third summand is our term. Rearrange this to get the desired result. *Q.E.D.*

PROOF OF PROPOSITION 3:

$$\frac{d^2 \log Y}{d \log A_j d \log A_i} = \frac{d \lambda_i}{d \log A_j}. \quad (18)$$

By definition,

$$\frac{d \log \lambda_i}{d \log A_j} = \left(\frac{1}{\rho_{ji}} - 1 \right) + \frac{d \log \lambda_j}{d \log A_j}, \quad (19)$$

which simplifies to

$$\frac{d \lambda_i}{d \log A_j} = \lambda_i \left(\frac{1}{\rho_{ji}} - 1 \right) + \frac{\lambda_i}{\lambda_j} \frac{d \lambda_j}{d \log A_j}. \quad (20)$$

Now apply Theorem 2 to the second summand to obtain the desired result. *Q.E.D.*

PROOF OF PROPOSITION 5: The allocation for labor is

$$\frac{l_{ig}}{\bar{l}_{ig}} = \left(\frac{\omega_{0i} A_i^{\frac{\theta-1}{\theta}}}{\sum \omega_{0j} A_j^{\frac{\theta-1}{\theta}}} \right)^{\frac{1}{1-\omega_g \frac{\theta-1}{\theta}}}.$$

Substituting this into the utility function gives

$$\frac{Y}{\bar{Y}} = \left(\sum_i \omega_{0i}^{\frac{\theta}{\theta(1-\omega_g)+\omega_g}} A_i^{\frac{\theta-1}{\theta(1-\omega_g)+\omega_g}} \right)^{\frac{\theta(1-\omega_g)+\omega_g}{\theta-1}}.$$

Then, for this economy,

$$\rho_{ji} = \rho = \frac{\theta(1-\omega_g) + \omega_g}{\theta(1-\omega_g) + \omega_g + (1-\theta)},$$

where $\omega_g = 0$ corresponds to

$$\rho = \theta,$$

which is the same as the no-reallocation case. On the other hand, for $\omega_g = 1$,

$$\rho = \frac{1}{2-\theta},$$

which is the same as the fully reallocation case. Note that this explodes when $\theta \geq 2$. For $\rho \in (0, 1)$, we get something in between the perfectly reallocation and no-reallocation special cases. *Q.E.D.*

PROOF OF PROPOSITION 6: Consumption is given by

$$Y = A \bar{Y} \left(\bar{a} \left(\frac{l}{\bar{l}} \right)^{\frac{\theta-1}{\theta}} + (1-\bar{a}) \left(\frac{X}{\bar{X}} \right)^{\frac{\theta-1}{\theta}} \right)^{\frac{\theta}{\theta-1}} - X.$$

The first-order condition gives

$$\frac{X}{\bar{X}} = (\bar{Y} A)^{\theta-1} (1-\bar{a})^{\theta} \bar{X}^{-\theta} Y.$$

Substituting this into the production function gives

$$Y = \frac{A\bar{Y}\bar{a}^{\frac{\theta}{\theta-1}}}{(1 - (1 - \bar{a})^\theta (\bar{Y}A/\bar{X})^{\theta-1})^{\frac{\theta}{\theta-1}}}.$$

This means that

$$Y = \frac{A\bar{Y}\bar{a}^{\frac{\theta}{\theta-1}}}{(1 - (1 - \bar{a})^\theta (\bar{Y}A/\bar{X})^{\theta-1})^{\frac{1}{\theta-1}}}.$$

Finally, note that

$$\frac{d \log Y}{d \log A} = \xi. \quad Q.E.D.$$

PROOF OF PROPOSITION 7: This follows as a special case of Proposition 9. Q.E.D.

PROOF OF COROLLARY 1: The proof is similar to that of Corollary 2. Q.E.D.

PROOF OF PROPOSITION 8: Denote the i th standard basis vector by e_i . Then, by assumption, $\Omega e_i = \Omega e_j$ and $b'e_i = b'e_j$. Repeated multiplication implies that $\Omega^n e_i = \Omega^n e_j$. This then implies that $\Psi e_i = \Psi e_j$ so that, in steady state, $\lambda_i = b'\Psi e_i = b'\Psi e_j = \lambda_j$. So the first-order impact of a shock is the same. Furthermore, substitution into (7) shows that the second-order impact of a shock is also the same. Q.E.D.

PROOF OF PROPOSITION 9: Denote the $N \times F$ matrix corresponding to Ω_{if} by α_{if} . By Shephard's lemma,

$$\frac{d \log p_i}{d \log A_k} = -\mathbf{1}(i=k) + \sum_j \Omega_{ij} \frac{d \log p_j}{d \log A_k} + \sum_f \alpha_{if} \frac{d \log w_f}{d \log A_k}. \quad (21)$$

Invert this system to get

$$\frac{d \log p_i}{d \log A_k} = -\Psi_{ik} + \sum_f \Psi_{if} \frac{d \log w_f}{d \log A_k}, \quad (22)$$

where $\Psi_{(f)} = (I - \Omega)^{-1} \alpha_{(f)}$. Note that $b'\Psi_f = \Lambda_f$.

Denote the household's final demand expenditure share Ω_{0i} by b_i . Then, for a factor L , we have

$$\begin{aligned} \frac{d \Lambda_f}{d \log A_k} &= \sum_i b_i (1 - \theta_0) \left[-\Psi_{ik} + \sum_g \Psi_{ig} \frac{d \log w_g}{d \log A_k} \right] \Psi_{if} \\ &\quad + \sum_j (1 - \theta_j) \lambda_j \sum_i \Omega_{ji} \left[-\Psi_{ik} + \sum_g \Psi_{ig} \frac{d \log w_g}{d \log A_k} + \Psi_{jk} - \sum_g \Psi_{jg} \frac{d \log w_g}{d \log A_k} \right] \Psi_{if} \\ &\quad + (\theta_k - 1) \lambda_k \sum_i \Omega_{ki} \Psi_{if}. \end{aligned}$$

Simplify this to

$$\begin{aligned}
 d\Lambda_f &= (\theta_0 - 1) \left(\sum_i b_i \Psi_{ik} \Psi_{if} - \sum_i b_i \Psi_{if} \sum_g \Psi_{ig} \frac{d \log w_g}{d \log A_k} \right) \\
 &\quad + \sum_j (\theta_j - 1) \lambda_j \left[\sum_i \Omega_{ji} \Psi_{ik} \Psi_{if} - \sum_i \Omega_{ji} \Psi_{jk} \Psi_{if} \right] \\
 &\quad + \sum_j (1 - \theta_j) \lambda_j \left[\sum_i \Omega_{ji} \sum_g (\Psi_{ig} - \Psi_{jg}) \frac{d \log w_g}{d \log A_k} \Psi_{if} \right] \\
 &\quad + (\theta_k - 1) \lambda_k \sum_i \Omega_{ki} \Psi_{if} \\
 &= (\theta_0 - 1) \left(\sum_i b_i \Psi_{ik} \Psi_{if} - \sum_i b_i \Psi_{if} \sum_g \Psi_{ig} \frac{d \log w_g}{d \log A_k} \right) \\
 &\quad + \sum_j (\theta_j - 1) \lambda_j \left[\sum_i \Omega_{ji} \Psi_{ik} \Psi_{if} - \left(\sum_i \Omega_{ji} \Psi_{if} \right) \left(\sum_i \Omega_{ji} \Psi_{ik} \right) \right] \\
 &\quad + \sum_j (1 - \theta_j) \lambda_j \left[\sum_i \Omega_{ji} \sum_g (\Psi_{ig} - \Psi_{jg}) \frac{d \log w_g}{d \log A_k} \Psi_{if} \right] \\
 &= (\theta_0 - 1) \left(\sum_i b_i \Psi_{ik} \Psi_{if} - \sum_i b_i \Psi_{if} \sum_g \Psi_{ig} \frac{d \log w_g}{d \log A_k} \right) \\
 &\quad + \sum_j (\theta_j - 1) \lambda_j \text{Cov}_{\Omega^{(j)}}(\Psi_{(k)}, \Psi_{(f)}) \\
 &\quad + \sum_j (1 - \theta_j) \lambda_j \sum_g \left[\sum_i \Omega_{ji} \Psi_{ig} \Psi_{if} - \left(\sum_i \Omega_{ji} \Psi_{if} \right) \left(\sum_i \Omega_{ji} \Psi_{ig} \right) \right] \frac{d \log w_g}{d \log A_k} \\
 &= (\theta_0 - 1) \left(\sum_i b_i \Psi_{ik} \Psi_{if} - \sum_i b_i \Psi_{if} \sum_g \Psi_{ig} \frac{d \log w_g}{d \log A_k} \right) \\
 &\quad + \sum_j (\theta_j - 1) \lambda_j \text{Cov}_{\Omega^{(j)}}(\Psi_{(k)}, \Psi_{(f)}) \\
 &\quad + \sum_j (1 - \theta_j) \lambda_j \sum_g \text{Cov}_{\Omega^{(j)}}(\Psi_{(g)}, \Psi_{(f)}) \frac{d \log w_g}{d \log A_k} \\
 &= (\theta_0 - 1) \left(\sum_i b_i \Psi_{ik} \Psi_{if} - \sum_i b_i \Psi_{if} \sum_g \Psi_{ig} \frac{d \log w_g}{d \log A_k} \right) \\
 &\quad + \sum_j (\theta_j - 1) \lambda_j \text{Cov}_{\Omega^{(j)}} \left(\Psi_{(k)} - \sum_g \Psi_{(g)} \frac{d \log w_g}{d \log A_k}, \Psi_{(f)} \right) \\
 &= (\theta_0 - 1) \left(\sum_i b_i \Psi_{if} \left(\Psi_{ik} - \sum_g \Psi_{ig} \frac{d \log w_g}{d \log A_k} \right) \right)
 \end{aligned}$$

$$\begin{aligned}
& + \sum_j (\theta_j - 1) \lambda_j \text{Cov}_{\Omega(j)} \left(\Psi_{(k)} - \sum_g \Psi_{(g)} \frac{d \log w_g}{d \log A_k}, \Psi_{(f)} \right) \\
& = (\theta_0 - 1) \text{Cov}_b \left(\Psi_{(k)} - \sum_g \Psi_{(g)} \frac{d \log w_g}{d \log A_k}, \Psi_{(f)} \right) \\
& + (\theta_0 - 1) \left(\lambda_k - \sum_g \Lambda_g \frac{d \log w_g}{d \log A_k} \right) \lambda_f \\
& + \sum_j (\theta_j - 1) \lambda_j \text{Cov}_{\Omega(j)} \left(\Psi_{(k)} - \sum_g \Psi_{(g)} \frac{d \log w_g}{d \log A_k}, \Psi_{(f)} \right).
\end{aligned}$$

Hence, for a productivity shock $d \log A_k$, letting Λ_f be demand for factor f , we have

$$\begin{aligned}
\frac{d \Lambda_f}{d \log A_k} & = (\theta_0 - 1) \text{Cov}_b \left(\Psi_{(k)} - \sum_g \Psi_{(g)} \frac{d \log w_g}{d \log A_k}, \Psi_{(f)} \right) \\
& + \sum_j (\theta_j - 1) \lambda_j \text{Cov}_{\Omega(j)} \left(\Psi_{(k)} - \sum_g \Psi_{(g)} \frac{d \log w_g}{d \log A_k}, \Psi_{(f)} \right) \\
& + (\theta_0 - 1) \left(\lambda_k - \sum_g \Lambda_g \frac{d \log w_g}{d \log A_k} \right) \Lambda_f.
\end{aligned} \tag{23}$$

We have

$$\frac{d \log w_f}{d \log A_k} = \frac{d \log Y}{d \log A_k} + \frac{1}{\Lambda_f} \frac{d \Lambda_f}{d \log A_k} = \lambda_k + \frac{1}{\Lambda_f} \frac{d \Lambda_f}{d \log A_k}. \tag{24}$$

Substituting this expression back into the formula, we get

$$\begin{aligned}
\frac{d \log \Lambda_f}{d \log A_k} & = (\theta_0 - 1) \frac{1}{\Lambda_f} \text{Cov}_b \left(\Psi_{(k)} - \sum_g \Psi_{(g)} d \log \Lambda_g, \Psi_{(f)} \right) \\
& + \sum_j (\theta_j - 1) \frac{\lambda_j}{\Lambda_f} \text{Cov}_{\Omega(j)} \left(\Psi_{(k)} - \sum_g \Psi_{(g)} \frac{d \log \Lambda_g}{d \log A_k}, \Psi_{(f)} \right).
\end{aligned} \tag{25}$$

The proof obtains by labeling final demand as producer 0. The derivation of the expression for $d \log \lambda_i / d \log A_k$ is similar. *Q.E.D.*

PROOF OF COROLLARY 2: Note that since $\theta_i = \theta$ for every i , market clearing for a good i (neglecting the normalizing constants and setting the household's price index to be the numeraire) is

$$p_i^\theta y_i = \omega_{0i}^\theta Y + \sum_j p_j^\theta y_j \omega_{ji}^\theta. \tag{26}$$

Hence, letting $p^\theta y$ denote the vector whose i th element is $p_i^\theta y_i$ and $b_i = \omega_{0i}$, we can write

$$p^\theta y = b' (I - \omega^{\circ\theta})^{-1} Y, \tag{27}$$

where $\omega^{\circ\theta}$ is the matrix of ω_{ij} raised to θ elementwise. Let $\tilde{b}' = b'(I - \omega^{\circ\theta})^{-1}$. This is reminiscent of the supplier centrality defined by Baqaee (2018).

Furthermore, market clearing for labor type k is

$$w_k^\theta \bar{l}_k = A_k^{-1} \sum_i p_i^\theta y_i \omega_{ik}^\theta, \quad (28)$$

where we use the fact that productivity shocks affect only the stock of factors. Rearrange this to get

$$w_k^\theta \bar{l}_k = A_k^{-1} Y \left(\sum_i \tilde{b}_i \omega_{ik}^\theta \right), \quad (29)$$

whence

$$w_k = \left(\frac{Y}{A_k \bar{l}_k} \right)^{\frac{1}{\theta}} \left(\sum_i \tilde{b}_i \omega_{ik}^\theta \right)^{\frac{1}{\theta}}. \quad (30)$$

To complete the proof, note that

$$Y = \sum_k w_k A_k \bar{l}_k = \left(\frac{Y}{A_k \bar{l}_k} \right)^{\frac{1}{\theta}} \left(\sum_i \tilde{b}_i \omega_{ik}^\theta \right)^{\frac{1}{\theta}} A_k \bar{l}_k. \quad (31)$$

Rearrange this to get a closed-form expression for output

$$Y = \left(\sum_k (A_k \bar{l}_k)^{\frac{\theta-1}{\theta}} \left(\sum_i \tilde{b}_i \omega_{ik}^\theta \right)^{\frac{1}{\theta}} \right)^{\frac{\theta}{\theta-1}}. \quad (32)$$

Since Y can be written in closed form as a CES aggregate of the underlying productivity shocks, Corollary 2 follows immediately. Q.E.D.

PROOF OF PROPOSITION 10: By Lemma (5.8) from Theil (1967, p. 222), we know that

$$\begin{aligned} & \log(Y(A + \Delta)/Y(A)) \\ &= \frac{1}{2} [\nabla \log Y(A + \Delta) + \nabla \log Y(A)]' [\log(A + \Delta) - \log(A)] + O(\Delta^3). \end{aligned} \quad (33)$$

Hulten (1978) then implies that $\nabla \log Y(A) = \lambda(A)$ and $\nabla \log Y(A + \Delta) = \lambda(A + \Delta)$. Q.E.D.

REFERENCES

- ACEMOGLU, D., U. AKCIGIT, AND W. KERR (2015): "Networks and the Macroeconomy: An Empirical Exploration," NBER Macro Conference. [1179]
- ACEMOGLU, D., D. AUTOR, D. DORN, G. H. HANSON, AND B. PRICE (2016): "Import Competition and the Great US Employment Sag of the 2000s," *Journal of Labor Economics*, 34(S1), S141–S198. [1185]
- ACEMOGLU, D., V. M. CARVALHO, A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80 (5), 1977–2016. [1158,1162,1163]
- ACEMOGLU, D., A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2015): "Systemic Risk and Stability in Financial Networks," *The American Economic Review*, 105 (2), 564–608. [1157]

- (2016): “Networks, Shocks, and Systemic Risk,” in *The Oxford Handbook of the Economics of Networks*. [1157,1175]
- (2017): “Microeconomic Origins of Macroeconomic Tail Risks,” *The American Economic Review*, 107 (1), 54–108. [1158,1163,1188]
- AGHION, P., AND P. W. HOWITT (2008): *The Economics of Growth*. MIT Press. [1156]
- ALTINOGLU, L. (2016): “The Origins of Aggregate Fluctuations in a Credit Network Economy”. [1158]
- ARKOLAKIS, C., A. COSTINOT, AND A. RODRÍGUEZ-CLARE (2012): “New Trade Models, Same Old Gains?” *American Economic Review*, 102 (1), 94–130. [1175]
- ATALAY, E. (2017): “How Important Are Sectoral Shocks?” *American Economic Journal: Macroeconomics*, 9 (4), 254–280. [1157,1158,1186]
- AUTOR, D. H., D. DORN, AND G. H. HANSON (2016): “The China Shock: Learning From Labor-Market Adjustment to Large Changes in Trade,” *Annual Review of Economics*, 8, 205–240. [1185]
- BALLESTER, C., A. CALVÓ-ARMENGOL, AND Y. ZENOU (2006): “Who’s Who in Networks. Wanted: The Key Player,” *Econometrica*, 74 (5), 1403–1417. [1157]
- BAQAEE, D., AND E. FARHI (2018a): “The Microeconomic Foundations of Aggregate Production Functions,” Technical Report, National Bureau of Economic Research. [1181,1194]
- BAQAEE, D. R., AND E. FARHI (2019): “Supplement to ‘The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten’s Theorem’,” *Econometrica Supplemental Material*, 87, <https://doi.org/10.3982/ECTA15202>. [1165]
- BAQAEE, D. R. (2015): “Targeted fiscal policy.” [1174]
- (2018): “Cascading Failures in Production Networks,” *Econometrica*, 86 (5), 1819–1838. [1158,1176,1178,1201]
- BAQAEE, D. R., AND E. FARHI (2018b): “Macroeconomics With Heterogeneous Agents and Input–Output Networks,” Technical Report, NBER Working Papers 24684, National Bureau of Economic Research, Inc. [1158,1194]
- (2018c): “Productivity and Misallocation in General Equilibrium,” Technical Report, National Bureau of Economic Research. [1158,1194]
- (2019): “Networks, Barriers, and Trade,” Technical Report. [1175,1194]
- BARRO, R. J. (2006): “Rare Disasters and Asset Markets in the Twentieth Century,” *The Quarterly Journal of Economics*, 823–866. [1188]
- BARROT, J.-N., AND J. SAUVAGNAT (2016): “Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks,” *The Quarterly Journal of Economics*. [1157]
- BARTELME, D., AND Y. GORODNICHENKO (2015): “Linkages and Economic Development,” Technical Report, National Bureau of Economic Research. [1158,1163]
- BAUMOL, W. J. (1967): “Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis,” *The American Economic Review*, 57 (3), 415–426. [1157,1192]
- BIGIO, S., AND J. LAO (2016): “Financial Frictions in Production Networks,” Technical Report. [1158,1163]
- BLACKORBY, C., AND R. R. RUSSELL (1989): “Will the Real Elasticity of Substitution Please Stand up? (A Comparison of the Allen/Uzawa and Morishima Elasticities),” *The American Economic Review*, 79 (4), 882–888. [1161]
- BOEHM, C. E., A. FLAAEN, AND N. PANDALAI-NAYAR (2017): “Input Linkages and the Transmission of Shocks: Firm-Level Evidence From the 2011 Tohoku Earthquake,” *The Review of Economics and Statistics*. [1157,1186]
- BURSTEIN, A., AND J. CRAVINO (2015): “Measured Aggregate Gains From International Trade,” *American Economic Journal: Macroeconomics*, 7 (2), 181–218. [1158]
- CARVALHO, V., AND X. GABAIX (2013): “The Great Diversification and Its Undoing,” *The American Economic Review*, 103 (5), 1697–1727. [1158,1185,1187]
- CARVALHO, V. M. (2010): “Aggregate Fluctuations and the Network Structure of Intersectoral Trade”. [1158]
- COSTINOT, A., AND A. RODRIGUEZ-CLARE (2014): “Trade Theory With Numbers: Quantifying the Consequences of Globalization,” *Handbook of International Economics*, 4, 197. [1162]
- DI GIOVANNI, J., A. A. LEVCHENKO, AND I. MÉJEAN (2014): “Firms, Destinations, and Aggregate Fluctuations,” *Econometrica*, 82 (4), 1303–1340. [1158]
- DURLAUF, S. N. (1993): “Nonergodic Economic Growth,” *The Review of Economic Studies*, 60 (2), 349–366. [1157]
- ELLIOTT, M., B. GOLUB, M. O. JACKSON et al. (2014): “Financial Networks and Contagion,” *American Economic Review*, 104 (10), 3115–3153. [1157]
- FEENSTRA, R. C., AND G. H. HANSON (1996): “Globalization, Outsourcing, and Wage Inequality,” *The American Economic Review*, 86 (2), 240. [1162]

- FOERSTER, A. T., P.-D. G. SARTE, AND M. W. WATSON (2011): "Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, 119 (1), 1–38. [1158]
- GABAIX, X. (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79 (3), 733–772. [1158, 1162, 1187]
- (2016): "Power Laws in Economics: An Introduction," *Journal of Economic Perspectives*, 30 (1), 185–206. [1158]
- GOMME, P., AND P. RUPERT (2007): "Theory, Measurement and Calibration of Macroeconomic Models," *Journal of Monetary Economics*, 54 (2), 460–497. [1156]
- GRASSI, B. (2017): "IO in I-o: Competition and Volatility in Input–Output Networks," Technical Report. [1158]
- GROSSMAN, G. M., E. HELPMAN, E. OBERFIELD, AND T. SAMPSON (2016): "Balanced Growth Despite Uzawa," Technical Report, National Bureau of Economic Research. [1156]
- HAMILTON, J. D. (2003): "What Is an Oil Shock?" *Journal of Econometrics*, 113 (2), 363–398. [1190]
- (2013): "Historical Oil Shocks," in *Routledge Handbook of Major Events in Economic History*, ed. by R. E. Parker and R. M. Whaples. Routledge, 239–265. Chapter 21. [1192]
- HERRENDORF, B., R. ROGERSON, AND A. VALENTINYI (2013): "Two Perspectives on Preferences and Structural Transformation," *American Economic Review*, 103 (7), 2752–2789. [1186, 1193]
- HICKS, J. (1932): *Theory of Wages*. Springer. [1161]
- HORVATH, M. (1998): "Cyclical and Sectoral Linkages: Aggregate Fluctuations From Independent Sectoral Shocks," *Review of Economic Dynamics*, 1 (4), 781–808. [1158]
- (2000): "Sectoral Shocks and Aggregate Fluctuations," *Journal of Monetary Economics*, 45 (1), 69–106. [1158]
- HULTEN, C. R. (1978): "Growth Accounting With Intermediate Inputs," *The Review of Economic Studies*, 511–518. [1155, 1160, 1184, 1201]
- JONES, C. I. (2011): "Intermediate Goods and Weak Links in the Theory of Economic Development," *American Economic Journal: Macroeconomics*, 1–28. [1158, 1162, 1171]
- (2013): "Input–Output Economics," in *Advances in Economics and Econometrics: Tenth World Congress*, Vol. 2. Cambridge University Press, 419. [1158]
- JORGENSEN, D. W., F. GOLLOP, AND B. M. FRAUMENI (1987): "Productivity and US Economic Growth." [1185]
- JOVANOVIC, B. (1987): "Micro Shocks and Aggregate Risk," *The Quarterly Journal of Economics*, 395–409. [1157]
- KIM, S.-J., H. S. SHIN et al. (2013): "Working Capital, Trade and Macro Fluctuations," Technical Report. [1162]
- LIU, E. (2017): "Industrial Policies and Economic Development," Technical Report. [1158]
- LONG, J. B., AND C. I. PLOSSER (1983): "Real Business Cycles," *The Journal of Political Economy*, 39–69. [1156, 1158, 1163]
- LUCAS, R. E. (1987): *Models of Business Cycles*, Vol. 26. Oxford: Basil Blackwell. [1155, 1157, 1166, 1188]
- MORISHIMA, M. (1967): "A few Suggestions on the Theory of Elasticity," *Keizai Hyoron (Economic Review)*, 16, 144–150. [1161]
- NOTOWIDIGDO, M. J. (2011): "The Incidence of Local Labor Demand Shocks," Technical Report, National Bureau of Economic Research. [1185]
- OBERFIELD, E., AND D. RAVAL (2014): "Micro Data and Macro Technology," Technical Report, National Bureau of Economic Research. [1186]
- SCHEINKMAN, J. A., AND M. WOODFORD (1994): "Self-Organized Criticality and Economic Fluctuations," *The American Economic Review*, 417–421. [1157]
- STELLA, A. (2015): "Firm Dynamics and the Origins of Aggregate Fluctuations," *Journal of Economic Dynamics and Control*, 55, 71–88. [1158]
- THEIL, H. (1967): "Economics and Information Theory," Technical Report. [1201]
- TÖRNOVIST, L. (1936): "The Bank of Finland's Consumption Price Index," *Bank of Finland Monthly Bulletin*, 10. [1191]

Co-editor Giovanni L. Violante handled this manuscript.

Manuscript received 23 March, 2017; final version accepted 22 February, 2019; available online 28 February, 2019.