Online Appendix to Productivity and Misallocation in General Equilibrium

David Rezza Baqaee Emmanuel Farhi

March 10, 2020

*Emails: baqaee@econ.ucla.edu, efarhi@harvard.edu. We thank Philippe Aghion, Pol Antras, Andrew Atkeson, Susanto Basu, John Geanakoplos, Ben Golub, Gita Gopinath, Dale Jorgenson, Marc Melitz, Ben Moll, Matthew Shapiro, Dan Trefler, Venky Venkateswaran and Jaume Ventura for their valuable comments. We are especially grateful to Natalie Bau for detailed conversations. We thank German Gutierrez, Thomas Philippon, Jan De Loecker, and Jan Eeckhout for sharing their data. We thank Thomas Brzustowski and Maria Voronina for excellent research assistance.

Appendix B Empirical Comparison with Petrin-Levinsohn

Figure 8 constructs the Petrin-Levinsohn decomposition with markups obtained from the production-function approach, at the firm level. The factor-growth-miscounting term is introduced to correct for the fact that the Petrin-Levinsohn decomposition applies to the Solow residual, whereas ours applies to the distortion-adjusted Solow residual, where both residuals weigh the growth of each factor differently. It does not affect the pure technology and changes in allocative efficiency effects constructed using their procedure.

Figures 6a and 8 allow us to compare the different results that are obtained using our decomposition and the Petrin-Levinsohn decomposition. Compared to ours, the Petrin-Levinsohn decomposition finds lower contributions both for pure technology and for allocative efficiency. The different weights used to weigh labor and capital growth in the Solow residual vs. the distortion-adjusted Solow also lead to a sizable difference between the Solow residual and the distortion-adjusted Solow residual. The cumulated Solow residual is significantly lower than the cumulated distortion-adjusted Solow residual, and this is reflected in a sizable positive contribution of factor-growth miscounting.

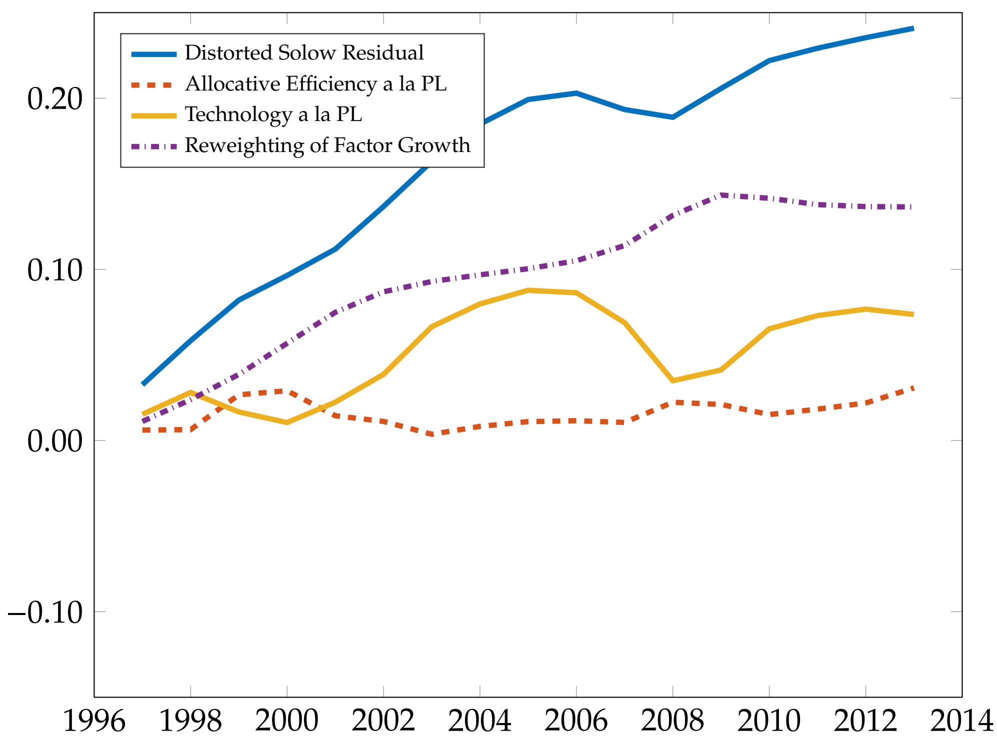

Figure 8: Petrin-Levinsohn decomposition of changes in aggregate TFP into pure changes in technology, changes in allocative efficiency, and factor under-counting, with markups obtained from the production-function approach, at the firm level.

Appendix C Data

We have two principal datasources: (i) aggregate data from the BEA, including the input-output tables and the national income and product accounts; (ii) firm-level data from Compustat. Below we describe how we treat the input-output data, merge it with firm-level estimates of markups, and how we estimate markups at the firm-level.

C.1 Input-Output and Aggregate Data

Our measure of real GDP growth, and growth in real factor quantities (labor and capital) come from the San Francisco Federal Reserve's dataset on total factor productivity.[1] Specifically, we use the variable "dY" for real GDP growth, "dK" for real capital growth, and "dLQ+dhours" for labor input growth.

Our input-output data comes from the BEA's annual input-output tables. We calibrate the data to the use tables from 1997-2015 before redefinitions. We also ignore the distinction between commodities and industries, assuming that each industry produces one commodity. For each year, this gives us the revenue-based expenditure share matrix Ω as well as the final demand budget shares b. We drop the government, scrap, and noncomparable imports sectors from our dataset, leaving us with 66 industries. We define the gross-operating surplus of each industry to be the residual from sales minus intermediate input costs and compensation of employees. The expenditures on capital, at the industry level, are equal to the gross operating surplus minus the share of profits (how we calculate the profit share is described shortly). If this number is negative, we set it equal to zero. If any value in Ω is negative, we set it to zero.

We have three sources of markup data. For each markup series, we compute the profit share (amongst Compustat firms) for each industry and year, and then we use that profit share to separate payments to capital from gross operating surplus in the BEA data for that industry and year. Conditional on the harmonic average of markups in each industry-year, we can recover the cost-based $\tilde{\Omega} = \mu\Omega$. If for an industry and year we do not observe any Compustat firms, then we assume that the profit share (and the average markup) of that industry is equal to the aggregate profit share (and the industry-level markup is the same as the aggregate markup).

We assume that the economy has an industry structure along the lines of Appendix H.4, so that all producers in each industry have the same production function up to a Hicks-neutral productivity shifter. This means that for each producer i and j in the same

[1] Available at https://www.frbsf.org/economic-research/indicators-data/total-factor-productivity-tfp/

industry $\tilde{\Omega}_{ik} = \tilde{\Omega}_{jk}$. To populate each industry with individual firms, we divide the sales of each industry across the firms in Compustat according to the sales share of these firms in Compustat. In other words, if some firm i's markup is μ_i and share of industry sales in Compustat is x, then we assume that the mass of firms in that industry whose markups are equal to μ_i is also equal to x. These assumptions allow us to use the markup data and market share information from Compustat, and the industry-level IO matrix from the BEA, to construct the firm-level cost-based IO matrix.

C.2 Estimates of Markups

Now, we briefly describe how our firm-level markup data is constructed. Firm-level data is from Compustat, which includes all public firms in the U.S. The database covers 1950 to 2016, but we restrict ourselves to post-1997 data since that is the start of the annual BEA data. We exclude firm-year observations with assets less than 10 million, with negative book or market value, or with missing year, assets, or book liabilities. We exclude firms with BEA code 999 because there is no BEA depreciation available for them; and Financials (SIC codes 6000-6999 or NAICS3 codes 520-525). Firms are mapped to BEA industry segments using 'Level 3' NAICS codes, according to the correspondence tables provided by the BEA. When NAICS codes are not available, firms are mapped to the most common NAICS category among those firms that share the same SIC code and have NAICS codes available.

C.2.1 Accounting Profits Approach

For the accounting-profit approach markups, we use operating income before depreciation, minus depreciation to arrive at accounting profits. Our measure of depreciation is the industry-level depreciation rate from the BEA's investment series. The BEA depreciation rates are better than the Compustat depreciation measures since accounting rules and tax incentives incentivize firms to depreciate assets too quickly. We use the expression

$$profits_i = \left(1 - \frac{1}{\mu_i}\right) sales_i,$$

to back out the markups for each firm in each year. We winsorize markups and changes in markups at the 5-95th percentile by year. Intuitively, this is equivalent to assuming that the cost of capital is simply the depreciation rate (equivalently, the risk-adjusted rate of return on capital is zero).

C.2.2 User Cost Approach

The user-cost approach markups are similar to the accounting profits but require a more careful accounting for the user cost of capital. For this measure, we rely on the replication files from Gutiérrez and Philippon (2016) provided German Gutierrez. For more information see Gutiérrez and Philippon (2016). To recover markups, we assume that operating surplus of each firm is equal to payments to both capital as well as economic rents due to markups. We write

$$OS_{i,t} = r_{k_i,t}K_{i,t} + \left(1 - \frac{1}{\mu_i}\right) sales_{i,t},$$

where $OS_{i,t}$ is the operating income of the firm after depreciation and minus income taxes, $r_{k_i,t}$ is the user-cost of capital and $K_{i,t}$ is the quantity of capital used by firm i in industry j in period t. This equation uses the fact that each firm has constant-returns to scale. In other words,

$$\frac{OS_{i,t}}{K_{i,t}} = r_{k_i,t} + \left(1 - \frac{1}{\mu_i}\right) \frac{sales_{i,t}}{K_{i,t}}, \tag{21}$$

To solve for the markup, we need to account for both the user cost (rental rate) of capital as well as the quantity of capital. The user-cost of capital is given by

$$r_{k_i,t} = r_t^s + KRP_j - (1 - \delta_{k_i,t})E(\Pi_{t+1}^k),$$

where r_t^s is the risk-free real rate, KPR_j is the industry-level capital risk premium, δ_j is the industry-level BEA depreciation rate, and $E(\Pi_{t+1}^k)$ is the expected growth in the relative price of capital. We assume that expected quantities are equal to the realized ones. To calculate the user-cost, the risk-free real rate is the yield on 10-year TIPS starting in 2003. Prior to 2003, we use the average spread between nominal and TIPS bonds to deduce the real rate from nominal bonds prior to 2003. KRP is computed using industry-level equity risk premia following Claus and Thomas (2001) using analyst forecasts of earnings from IBES and using current book value and the average industry payout ratio to forecast future book value. The depreciation rate is taken from BEA's industry-level depreciation rates. The capital gains $E(\Pi_{t+1}^k)$ is equal to the growth in the relative price of capital computed from the industry-specific investment price index relative to the PCE deflator. Finally, we use net property, plant, and equipment as the measure of the capital stock. This allows us to solve equation (21) for a time-varying firm-level measure of the markup. We winsorize markups and changes in markups at the 5-95th percentile by year.

C.2.3 Production Function Estimation Approach

For the production function estimation approach markups, we follow the procedure PF1 described by De Loecker et al. (2019) with some minor differences. We estimate the production function using Olley and Pakes (1996) rather than Levinsohn and Petrin (2003). We use CAPX as the instrument and COGS as a variable input. We use the classification based on SIC numbers instead of NAICS numbers since they are available for a larger fraction of the sample. Finally, we exclude firms with COGS-to-sales and XSGA-to-sales ratios in the top and bottom 2.5% of the corresponding year-specific distributions. As with the other series, we use Compustat excluding all firms that did not report SIC or NAICS indicators, and all firms with missing sales or COGS. Sales and COGS are deflated using the gross output price indices from KLEMS sector-level data. CAPX and PPEGT – using the capital price indices from the same source. Industry classification used in the estimation is based on the 2-digit codes whenever possible, and 1-digit codes if there are fewer than 500 observations for each industry and year.

To compute the PF Markups, we need to estimate elasticity of output with respect to variable inputs. This is because once we know the output-elasticity with respect to a variable input (in this case, the cost of goods sold or COGS), then following Hall (1988), the markup is

$$\mu_i = \frac{\partial \log F_i / \partial \log COGS_i}{\Omega_{i,COGS}},$$

where $\Omega_{i,COGS}$ is the firm's expenditures on COGS relative to its turnover.

The output-elasticities are estimated using Olley and Pakes (1996) methodology with the correction advocated by Ackerberg et al. (2015). To implement Olley-Pakes in Stata, we use the *prodest* Stata package. OP estimation requires:

(i) outcome variable: log sales,

(ii) "free" variable (variable inputs): log COGS,

(iii) "state" variable: log capital stock, measured as log PPEGT in the Compustat data,

(iv) "proxy" variable, used as an instrument for productivity: log investment, measured as log CAPX in Compustat data.

(v) in addition, SIC 3-digit and SIC 4-digit firm sales shares were used to control for markups .

Given these data, we run the estimation procedure for every sector and every year. Since panel data are required, we use 3-year rolling windows so that the elasticity estimates based on data in years $t-1$, t and $t+1$ are assigned to year t. The estimation procedure has two stages: in the first stage, log sales are regressed on the 3-rd degree polynomial of state, free, proxy and control variables in order to remove the measurement error and unanticipated shocks; in the second stage, we estimate elasticities of output with respect to variable inputs and the state variable by fitting an AR(1) process for productivity to the data (via GMM). Just like in De Loecker et al. (2019), we control for markups using a linear function of firm sales shares (sales share at the 4-digit industry level).

We use a Cobb-Douglas specification of industry production functions because of its simplicity and stability. This means that to be entirely internally consistent, in our structural counterfactual exercise regarding the effect of removing markups on aggregate TFP, we should focus on specifications with unitary elasticities across industries and factors. For example, the benchmark should now be the CD+CES specification in the second column of Table 2 instead of that in the first column. Imposing elasticities across industries and factors would only introduce minor quantitative differences as we navigate through the other columns, and would not change the corresponding quantitative conclusions much.

Appendix D Proofs

Throughout this appendix, we let the nominal GDP be the numeraire, so that $PY = \sum_{i=1}^{N} p_i c_i = 1$ or equivalently $d\log(\sum_{i=1}^{N} p_i c_i) = 0$. This numeraire is different from the GDP deflator defined such that the ideal price index of the household is unitary $P = 1$, or equivalently $d\log P = \sum_{i=1}^{N} b_i d\log p_i = 0$. A price p_i in the nominal GDP numeraire can easily be converted into a price $p_i Y$ in the GDP deflator numeraire, so that $d\log(p_i Y) = d\log p_i + d\log Y$.

Proof of Theorem 1. We start by proving some preliminary results. Let $\tilde{\Omega}^p$ be the $N \times N$ matrix corresponding to the first N rows and columns (corresponding to goods prices) of $\tilde{\Omega}$, so that $\tilde{\Omega}^p_{ij} = \tilde{\Omega}^p_{ij}$ for $(i, j) \in [1, N]^2$. Since $\tilde{\Omega}$ is block-diagonal over goods prices and factor prices, we have that for all $(i, j) \in [1, N]^2$,

$$[(I - \tilde{\Omega}^p)^{-1}]_{ij} = [(I - \tilde{\Omega})^{-1}]_{ij} = \tilde{\Psi}_{ij}. \tag{22}$$

In addition, using

$$1 = \sum_{j=1}^{N} \tilde{\Omega}_{ij} + \sum_{f=1}^{F} \tilde{\Omega}_{if}, \tag{23}$$

which we can rewrite as

$$1^p = \tilde{\Omega}^p 1^p + \tilde{\Omega} 1^f, \tag{24}$$

where 1^p is a $N \times 1$ vector of ones and 1^f is a $F \times 1$ vector of ones. This in turn implies that

$$1^p = (I - \tilde{\Omega}^p)^{-1} \tilde{\Omega} 1^f, \tag{25}$$

and hence

$$1^p = \tilde{\Psi} 1^f, \tag{26}$$

$$1 = b' \tilde{\Psi} 1^f, \tag{27}$$

and finally, using

$$b' \tilde{\Psi} = \tilde{\lambda}', \tag{28}$$

we get

$$1 = \sum_{f=1}^{F} \tilde{\Lambda}_f. \tag{29}$$

We now move on to the main proof. By Sheppard's lemma, we have

$$d \log p_i = -d \log A_i + d \log \mu_i + \sum_{j=1}^{N} \tilde{\Omega}_{ij} d \log p_j + \sum_{f=1}^{F} \tilde{\Omega}_{if} d \log w_f. \tag{30}$$

In the nominal GDP numeraire where $\sum p_i c_i = 1$, we have $w_f L_f = \Lambda_f$. Since we hold factor supplies fixed, we have

$$d \log w_f = d \log \Lambda_f. \tag{31}$$

This implies that

$$d \log p_i = -d \log A_i + d \log \mu_i + \sum_{j=1}^{N} \tilde{\Omega}_{ij} d \log p_j + \sum_{f=1}^{F} \tilde{\Omega}_{if} d \log \Lambda_f. \tag{32}$$

We can rewrite this as

$$d \log p_i = \sum_{k=1}^{N} [(I - \tilde{\Omega}^p)^{-1}]_{ik} (-d \log A_k + d \log \mu_k) + \sum_{f=1}^{F} \sum_{k=1}^{N} [(I - \tilde{\Omega}^p)^{-1}]_{ik} \tilde{\Omega}_{kf} d \log \Lambda_f. \tag{33}$$

This implies that

$$d\log p_i = \sum_{k=1}^{N} \tilde{\Psi}_{ik}(-d\log A_k + d\log \mu_k) + \sum_{f=1}^{F}\sum_{k=1}^{N} \tilde{\Psi}_{ik}\tilde{\Omega}_{kf} d\log \Lambda_f. \tag{34}$$

This in turn implies that

$$d\log p_i = \sum_{k=1}^{N} \tilde{\Psi}_{ik}(-d\log A_k + d\log \mu_k) + \sum_{f=1}^{F} \tilde{\Psi}_{if} d\log \Lambda_f. \tag{35}$$

This can be rewritten in vector form as

$$d\log p = \sum_{k=1}^{N} \tilde{\Psi}_{(k)}(-d\log A_k + d\log \mu_k) + \sum_{f=1}^{F} \tilde{\Psi}_{(f)} d\log \Lambda_f, \tag{36}$$

where $\tilde{\Psi}_{(k)}$ and $\tilde{\Psi}_{(f)}$ are the k-th and f-th columns of $\tilde{\Psi}$, respectively. Since

$$d\log Y = -b' d\log p = -\sum_{i=1}^{N} b_i d\log p_i, \tag{37}$$

and since

$$b'\tilde{\Psi} = \tilde{\lambda}', \tag{38}$$

we get finally get

$$d\log Y = \sum_{k=1}^{N} \tilde{\lambda}_k d\log A_k - \sum_{k=1}^{N} \tilde{\lambda}_k d\log \mu_k - \sum_{f=1}^{F} \tilde{\Lambda}_f d\log \Lambda_f. \tag{39}$$

which proves Theorem 1.

■

Proofs of Propositions 2 and 3. We have

$$d\Omega_{ji} = -\Omega_{ji} d\log \mu_j + \frac{1}{\mu_j}(\theta_j - 1)\left(d\log p_i - \sum_{l} \tilde{\Omega}_{jl} d\log p_l\right). \tag{40}$$

or equivalently

$$d\Omega_{ji} = -\Omega_{ji} d\log \mu_j + \frac{1}{\mu_j}(\theta_j - 1) Cov_{\tilde{\Omega}^{(j)}}(d\log p, I_{(i)}), \tag{41}$$

where $I_{(i)}$ is the ith column of the identity matrix I. Using

$$d\log p = -\sum_k \tilde{\Psi}_{(k)} d\log A_k + \sum_k \tilde{\Psi}_{(k)} d\log \mu_k + \sum_f \tilde{\Psi}_{(f)} d\log \Lambda_f, \tag{42}$$

we can rewrite this as

$$d\Omega_{ji} = -\Omega_{ji} d\log \mu_j + \frac{1}{\mu_j}(\theta_j - 1) Cov_{\tilde{\Omega}^{(j)}}(\sum_k \tilde{\Psi}_{(k)} d\log A_k - \sum_k \tilde{\Psi}_{(k)} d\log \mu_k - \sum_g \tilde{\Psi}_{(g)} d\log \Lambda_g, I_{(i)}), \tag{43}$$

Using $\Psi = (I - \Omega)^{-1}$, we get

$$d\Psi = \Psi d\Omega \Psi. \tag{44}$$

Combining, we get

$$\begin{aligned} d\Psi_{mn} &= -\sum_j \Psi_{mj} d\log \mu_j \sum_i \Omega_{ji} \Psi_{in} \\ &+ \sum_j \frac{\Psi_{mj}}{\mu_j}(\theta_j - 1) Cov_{\tilde{\Omega}^{(j)}}(\sum_k \tilde{\Psi}_{(k)} d\log A_k - \sum_k \tilde{\Psi}_{(k)} d\log \mu_k - \sum_g \tilde{\Psi}_{(g)} d\log \Lambda_g, \sum_i I_{(i)} \Psi_{in}). \end{aligned} \tag{45}$$

Using $\Omega\Psi = \Psi - I$, we can re-express this as

$$\begin{aligned} d\Psi_{mn} &= -\sum_j \Psi_{mj}(\Psi_{jn} - \delta_{jn}) d\log \mu_j \\ &+ \sum_j \frac{\Psi_{mj}}{\mu_j}(\theta_j - 1) Cov_{\tilde{\Omega}^{(j)}}(\sum_k \tilde{\Psi}_{(k)} d\log A_k - \sum_k \tilde{\Psi}_{(k)} d\log \mu_k - \sum_g \tilde{\Psi}_{(g)} d\log \Lambda_g, \Psi_{(n)}). \end{aligned} \tag{46}$$

Using $b'\Psi = \lambda$ in turn implies that

$$\begin{aligned} d\lambda_n &= -\sum_j \lambda_j(\Psi_{jn} - \delta_{jn}) d\log \mu_j \\ &+ \frac{\lambda_j}{\mu_j}(\theta_j - 1) Cov_{\tilde{\Omega}^{(j)}}(\sum_k \tilde{\Psi}_{(k)} d\log A_k - \sum_k \tilde{\Psi}_{(k)} d\log \mu_k - \sum_g \tilde{\Psi}_{(g)} d\log \Lambda_g, \Psi_{(n)}). \end{aligned} \tag{47}$$

Finally, dividing trough by λ_n, we get

$$d\log \lambda_n = -\sum_j \lambda_j \frac{\Psi_{jn} - \delta_{jn}}{\lambda_n} d\log \mu_j$$

$$+\frac{\lambda_j}{\mu_j}(\theta_j-1)Cov_{\tilde{\Omega}^{(j)}}(\sum_k \tilde{\Psi}_{(k)} d\log A_k - \sum_k \tilde{\Psi}_{(k)} d\log \mu_k - \sum_g \tilde{\Psi}_{(g)} d\log \Lambda_g, \frac{\Psi_{(n)}}{\lambda_n}). \quad (48)$$

Applying this to a factor share yields

$$d\log \Lambda_f = -\sum_j \lambda_j \frac{\Psi_{jf}}{\Lambda_f} d\log \mu_j$$
$$+\frac{\lambda_j}{\mu_j}(\theta_j-1)Cov_{\tilde{\Omega}^{(j)}}(\sum_k \tilde{\Psi}_{(k)} d\log A_k - \sum_k \tilde{\Psi}_{(k)} d\log \mu_k - \sum_g \tilde{\Psi}_{(g)} d\log \Lambda_g, \frac{\Psi_{(f)}}{\Lambda_f}). \quad (49)$$

Re-arranging the indices to make them consistent with the results stated in the main text, we get

$$d\log \lambda_i = -\sum_k \lambda_k \frac{\Psi_{ki}-\delta_{ki}}{\lambda_i} d\log \mu_k$$
$$+\frac{\lambda_j}{\mu_j}(\theta_j-1)Cov_{\tilde{\Omega}^{(j)}}(\sum_k \tilde{\Psi}_{(k)} d\log A_k - \sum_k \tilde{\Psi}_{(k)} d\log \mu_k - \sum_g \tilde{\Psi}_{(g)} d\log \Lambda_g, \frac{\Psi_{(i)}}{\lambda_i}). \quad (50)$$

Applying this to a factor share yields

$$d\log \Lambda_f = -\sum_k \lambda_k \frac{\Psi_{kf}}{\Lambda_f} d\log \mu_k$$
$$+\frac{\lambda_j}{\mu_j}(\theta_j-1)Cov_{\tilde{\Omega}^{(j)}}(\sum_k \tilde{\Psi}_{(k)} d\log A_k - \sum_k \tilde{\Psi}_{(k)} d\log \mu_k - \sum_g \tilde{\Psi}_{(g)} d\log \Lambda_g, \frac{\Psi_{(f)}}{\Lambda_f}). \quad (51)$$

■

Proof of Proposition 5. From Baqaee and Farhi (2019b), we know that the output losses can be expressed as

$$\mathcal{L} = -\frac{1}{2}\sum_l (d\log \mu_l)\lambda_l d\log y_l. \quad (52)$$

We have

$$d\log y_l = d\log \lambda_l - d\log p_l, \quad (53)$$

$$d\log p_l = \sum_f \Psi_{lf} d\log \Lambda_f + \sum_k \Psi_{lk} d\log \mu_k, \quad (54)$$

where, from Proposition 3

$$d\log\lambda_l = \sum_k (\delta_{lk} - \frac{\lambda_k}{\lambda_l}\Psi_{kl}) d\log\mu_k - \sum_j \frac{\lambda_j}{\lambda_l}(\theta_j - 1) Cov_{\Omega^{(j)}}(\sum_k \Psi_{(k)} d\log\mu_k + \sum_g \Psi_{(g)} d\log\Lambda_g, \Psi_{(l)}), \tag{55}$$

$$d\log\Lambda_f = -\sum_k \lambda_k \frac{\Psi_{kf}}{\Lambda_f} d\log\mu_k - \sum_j \lambda_j(\theta_j - 1) Cov_{\Omega^{(j)}}(\sum_k \Psi_{(k)} d\log\mu_k + \sum_g \Psi_{(g)} d\log\Lambda_g, \frac{\Psi_{(f)}}{\Lambda_f}). \tag{56}$$

We will now use these expressions to replace in formula for the second-order loss function. We get

$$\begin{aligned}\mathcal{L} = &-\frac{1}{2}\sum_l \sum_k (\frac{\delta_{lk}}{\lambda_k} - \frac{\Psi_{kl}}{\lambda_l} - \frac{\Psi_{lk}}{\lambda_k})\lambda_k\lambda_l d\log\mu_k d\log\mu_l + \frac{1}{2}\sum_l \lambda_l d\log\mu_l \sum_f \Psi_{lf} d\log\Lambda_f \\ &+ \frac{1}{2}\sum_l \sum_j (d\log\mu_l)\lambda_j(\theta_j - 1) Cov_{\Omega^{(j)}}(\sum_k \Psi_{(k)} d\log\mu_k + \sum_g \Psi_{(g)} d\log\Lambda_g, \Psi_{(l)}).\end{aligned}$$

We can rewrite this expression as

$$\mathcal{L} = \mathcal{L}_I + \mathcal{L}_X \tag{57}$$

where

$$\begin{aligned}\mathcal{L}_I = &\frac{1}{2}\sum_k \sum_l [\frac{\Psi_{kl} - \delta_{kl}}{\lambda_l} + \frac{\Psi_{lk} - \delta_{lk}}{\lambda_k} + \frac{\delta_{kl}}{\lambda_l} - 1]\lambda_k\lambda_l d\log\mu_k d\log\mu_l \\ &+ \frac{1}{2}\sum_k \sum_l \sum_j d\log\mu_k d\log\mu_l \lambda_j(\theta_j - 1) Cov_{\Omega^{(j)}}(\Psi_{(k)}, \Psi_{(l)}),\end{aligned}$$

$$\begin{aligned}\mathcal{L}_X = &\frac{1}{2}\sum_l \sum_f (\frac{\Psi_{lf}}{\Lambda_f} - 1)\lambda_l\Lambda_f d\log\mu_l d\log\Lambda_f \\ &+ \frac{1}{2}\sum_l \sum_g d\log\mu_l d\log\Lambda_g \sum_j \lambda_j(\theta_j - 1) Cov_{\Omega^{(j)}}(\Psi_{(g)}, \Psi_{(l)}),\end{aligned}$$

where $d\log\Lambda$ is given by the usual expression.[2] The proof is finished by use of the

[2] We have used the intermediate step

$$\mathcal{L}_X = \frac{1}{2}\sum_l \sum_k \lambda_k\lambda_l d\log\mu_k d\log\mu_l + \frac{1}{2}\sum_l \sum_f d\log\mu_l d\log\Lambda_f \lambda_l \Psi_{lf}$$

following lemma. ■

Lemma 2. *The following identity holds:*

$$\sum_j \lambda_j \mu_j^{-1} Cov_{\tilde{\Omega}^{(j)}}(\tilde{\Psi}_{(k)}, \Psi_{(l)}) = \lambda_l \lambda_k [\frac{\tilde{\Psi}_{lk} - \delta_{lk}}{\lambda_k} + \frac{\Psi_{kl} - \delta_{kl}}{\lambda_l} + \frac{\delta_{lk}}{\lambda_k} - \frac{\tilde{\lambda}_k}{\lambda_k}]. \tag{58}$$

This holds for inefficient economies with multiple factors and applies when k and l are goods or factors.

Proof. We have

$$\sum_j \lambda_j \mu_j^{-1} Cov_{\tilde{\Omega}^{(j)}}(\tilde{\Psi}_{(k)}, \Psi_{(l)}) =$$

$$\sum_j \lambda_j \mu_j^{-1} \left[\sum_m \tilde{\Omega}_{jm} \tilde{\Psi}_{mk} \Psi_{ml} - \left(\sum_m \tilde{\Omega}_{jm} \tilde{\Psi}_{mk} \right) \left(\sum_m \tilde{\Omega}_{jm} \Psi_{ml} \right) \right],$$

or

$$\sum_j \lambda_j \mu_j^{-1} Cov_{\tilde{\Omega}^{(j)}}(\tilde{\Psi}_{(k)}, \Psi_{(l)}) =$$

$$\sum_j \lambda_j \sum_m \Omega_{jm} \tilde{\Psi}_{mk} \Psi_{ml} - \sum_j \lambda_j \mu_j^{-1} \left(\sum_m \tilde{\Omega}_{jm} \tilde{\Psi}_{mk} \right) \left(\sum_m \tilde{\Omega}_{jm} \Psi_{ml} \right),$$

or

$$\sum_j \lambda_j \mu_j^{-1} Cov_{\tilde{\Omega}^{(j)}}(\tilde{\Psi}_{(k)}, \Psi_{(l)}) =$$

$$\sum_j \lambda_j \sum_m \Omega_{jm} \tilde{\Psi}_{mk} \Psi_{ml} - \sum_j \lambda_j \tilde{\Psi}_{jk} \Psi_{jl}$$
$$+ \sum_j \lambda_j \tilde{\Psi}_{jk} \Psi_{jl} - \sum_j \lambda_j \mu_j^{-1} \left(\sum_m \tilde{\Omega}_{jm} \tilde{\Psi}_{mk} \right) \left(\sum_m \tilde{\Omega}_{jm} \Psi_{ml} \right).$$

From the fact that

$$\sum_j \lambda_j \left[\Psi_{jk} \Psi_{jl} - \sum_m \Omega_{jm} \Psi_{mk} \Psi_{ml} \right] = \lambda_k \lambda_l. \tag{59}$$

$$+ \frac{1}{2} \sum_l \sum_g d \log \mu_l d \log \Lambda_g \sum_j \lambda_j (\theta_j - 1) Cov_{\Omega^{(j)}}(\Psi_{(g)}, \Psi_{(l)}).$$

the equation above can be simplified to

$$\sum_j \lambda_j \mu_j^{-1} Cov_{\tilde{\Omega}^{(j)}}(\tilde{\Psi}_{(k)}, \Psi_{(l)}) = -\tilde{\lambda}_k \lambda_l + \sum_j \lambda_j \tilde{\Psi}_{jk} \Psi_{jl} - \sum_j \lambda_j \left(\tilde{\Psi}_{jk} - \delta_{jk}\right)(\Psi_{jl} - \delta_{jl}), \tag{60}$$

and finally

$$\sum_j \lambda_j \mu_j^{-1} Cov_{\tilde{\Omega}^{(j)}}(\tilde{\Psi}_{(k)}, \Psi_{(l)}) = \lambda_l \lambda_k [\frac{\tilde{\Psi}_{lk} - \delta_{lk}}{\lambda_k} + \frac{\Psi_{kl} - \delta_{kl}}{\lambda_l} + \frac{\delta_{lk}}{\lambda_k} - \frac{\tilde{\lambda}_k}{\lambda_k}]. \tag{61}$$

■

Appendix E Basu-Fernald and Petrin-Levinsohn in a Simple Example

To compare our decomposition with that of Basu-Fernald and Petrin-Levinsohn, we consider the simple economy in Figure 9. There are two factors L_1 and L_2. There are two producers 1 and 2. Producer 2 produces linearly from factor L_2 with productivity A_2. It does not charge any markup $\mu_2 = 1$. Producer 1 uses the factor L_1 and output of producer 2 to produce according to a CES production function with steady-state revenue-based expenditure shares ω_{1L_1} and ω_{12}, and with elasticity of substitution θ_1 (this elasticity will not matter in the calculations below). It charges a markup $\mu_1 > 1$.

Because this economy is acyclic, there is a unique feasible allocation, and it is efficient. There is no misallocation, and there cannot be any change in allocative efficiency. Our decomposition gives

$$\frac{d \log Y}{d \log A_2} = \underbrace{\tilde{\lambda}_2}_{\Delta\text{Technology}} + \underbrace{0}_{\Delta\text{Allocative Efficiency}},$$

and the decompositions of Basu-Fernald and Petrin-Levinsohn both give

$$\frac{d \log Y}{d \log A_2} = \underbrace{\lambda_2}_{\Delta\text{Technology}} + \underbrace{\tilde{\lambda}_2 - \lambda_2}_{\Delta\text{Allocative Efficiency}},$$

where $\lambda_2 = \omega_{12}$, and $\tilde{\lambda}_2 = \mu_1 \omega_{12}$.Since $\tilde{\lambda}_2 = \mu_1 \lambda_2 > \lambda_2$, this immediately implies that while our decomposition does not detect any change in allocative efficiency, those of Basu-Fernald, and Petrin-Levinsohn do detect changes in allocative efficiency.

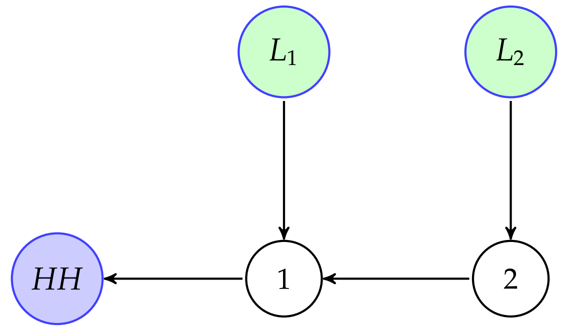

Figure 9: Acyclic economy where the solid arrows represent the flow of goods. The flow of profits and wages from firms to households has been suppressed in the diagram. The two factors in this economy are L_1 and L_2.

Appendix F Applying our Results with Endogenous Markups in a Simple Example

Consider the following endogenous-markup economy. There is a large number of industries indexed by $\mathcal{I}$. Within each industry, there is a finite number of producers $i \in \mathcal{I}$. Each producer produces linearly from labor with productivity A_i and charges a markup μ_i. The outputs of the different producers in each industry are combined into an industry output via a CES aggregator with elasticity $\theta_1 > 1$. The outputs of the different industries are combined into a final good via a Cobb-Douglas aggregator so that $\theta_0 = 1$. Following Atkeson and Burstein (2008), assume that producers play a static game of quantity competition. Specifically, each producer chooses its quantity taking as given the quantities chosen by the other producers as well as the wage and the price quantity of the final good. Under this assumption, producers do recognize that industry prices and quantities vary when that they change their quantities. This gives rise to endogenous markups i:

$$\frac{1}{\mu_i} = (1 - \frac{\lambda_i}{\lambda_{\mathcal{I}(i)}})(1 - \frac{1}{\theta_1}), \tag{62}$$

where $\mathcal{I}(i)$ is the industry of i, λ_i is its sales share, and $\lambda_{\mathcal{I}(i)}$ is the sales share of its industry. Hence, the markup of i is increasing in the relative sales share $\lambda_i/\lambda_{\mathcal{I}(i)}$ of i in its industry, and decreasing in the elasticity of substitution θ_1 across producers within an industry.

Suppose that all the industries are ex-ante identical. In each industry $\mathcal{I}$, there is a large producer k with $\lambda_k/\lambda_{\mathcal{I}} > 0$ and a continuum of atomistic producers, each with an infinitesimal relative sales share, but with strictly positive total relative sales share $1 - \lambda_k/\lambda_{\mathcal{I}(k)} > 0$. This implies that the markups of the atomistic producers are all constant at $1/(1 - 1/\theta_1)$.

Now consider a shock the productivity A_k of a single large producer k in a single industry $\mathcal{I}(k)$. The markup of producer i does not change if it is not in the industry of the shocked producer. The markup of an atomistic producer in the industry of the shocked producer does not change. And we can solve jointly for the change $\mathrm{d}\log\mu_k$ in the markup of producer k and for the change $\mathrm{d}\log\lambda_k$ in its sales share:

$$\mathrm{d}\log\mu_k = \frac{\lambda_k/\lambda_{\mathcal{I}(k)}}{1-\lambda_k/\lambda_{\mathcal{I}(k)}}\,\mathrm{d}\log\lambda_k, \quad \mathrm{d}\log\lambda_k = (\theta_1-1)(1-\lambda_k/\lambda_{\mathcal{I}(k)})(\mathrm{d}\log A_k - \mathrm{d}\log\mu_k), \tag{63}$$

where the first equation can be obtained by differentiating the markup equation (62), and where the second equation can be obtained by applying the propagation equations in Propositions 2 and 3 applied to producer k's sales share rather than to factor shares. This in turn implies that the markup μ_k of producer k increases endogenously with its productivity A_k according to

$$\frac{d\log\mu_k}{d\log A_k} = \frac{\frac{\lambda_k}{\lambda_{\mathcal{I}(k)}}(\theta_1-1)}{1+\frac{\lambda_k}{\lambda_{\mathcal{I}(k)}}(\theta_1-1)} > 0. \tag{64}$$

There is therefore imperfect pass-through of productivity shocks to prices. We then use the chain rule equation (16) with $Z = \log A_k$, in conjunction with the expressions for $\mathrm{d}\log\Lambda/\mathrm{d}\log A_k$ and for $\mathrm{d}\log\Lambda/\mathrm{d}\log\mu_k$ given by Propositions 2 and 3. We find that taking into account the endogenous change of the markups responsible for imperfect pass-through, the change $\mathrm{d}\log Y$ resulting from an increase $\mathrm{d}\log A_k > 0$ in the productivity of producer k is

$$\begin{aligned}\mathrm{d}\log Y &= \frac{d\log Y}{d\log A_k}\,\mathrm{d}\log A_k + \frac{\mathrm{d}\log Y}{\mathrm{d}\log\mu_k}\frac{\mathrm{d}\log\mu_k}{\mathrm{d}\log A_k}\,\mathrm{d}\log A_k \\ &= \lambda_k\left[1 + \frac{1-\frac{\lambda_k}{\lambda_{\mathcal{I}(k)}}}{1+\frac{\lambda_k}{\lambda_{\mathcal{I}(k)}}(\theta_1-1)}\frac{\frac{\lambda_i}{\lambda_{\mathcal{I}(k)}}(\theta_1-1)}{1+\frac{\lambda_k}{\lambda_{\mathcal{I}(k)}}}\right]\mathrm{d}\log A_k.\end{aligned}$$

If instead markups were exogenously fixed, we would have

$$\mathrm{d}\log Y = \frac{d\log Y}{d\log A_k}\,\mathrm{d}\log A_k = \lambda_k\left[1+\frac{\frac{\lambda_k}{\lambda_{\mathcal{I}(k)}}(\theta_1-1)}{1+\frac{\lambda_k}{\lambda_{\mathcal{I}(k)}}}\right]\mathrm{d}\log A_k,$$

which is strictly higher. We therefore see that imperfect pass-through via endogenous markups mitigates the impact of the shock.

Appendix G Standard-Form for Nested CES Economies

Throughout this section, variables with over-lines are normalizing constants equal to the values in steady-state. Since we are interested in log changes, the normalizing constants are irrelevant.[3]

Nested CES Economies in Standard Form

A CES economy in standard form is defined by a tuple (ω, θ, μ, F) and a set of normalizing constants $(\overline{y}, \overline{x})$. The $(N+F+1) \times (N+F+1)$ matrix ω is a matrix of input-output parameters where the first row and column correspond to the reproducible final good, the next N rows and columns correspond to reproducible goods and the last F rows and columns correspond to non-reproducible factors. The $(N+1) \times 1$ vector θ is a vector of microeconomic elasticities of substitution. Finally, the $N \times 1$ vector μ is a vector of markups/wedges for the N non-final reproducible goods.[4]

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

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

The other $N+1$ other goods are reproducible, and the production of a reproducible good k can be written as

$$\frac{y_k}{\overline{y}_k} = A_k \left(\sum_l \omega_{kl} \left(\frac{x_{kl}}{\overline{x}_{kl}} \right)^{\frac{\theta_k - 1}{\theta_k}} \right)^{\frac{\theta_k}{\theta_k - 1}},$$

where x_{lk} are intermediate inputs from l used by k. Each producer charges a markup over its marginal cost μ_k. Producer 0 represents final-demand and its production function the final-demand aggregator so that

$$\frac{Y}{\overline{Y}} = \frac{y_0}{\overline{y}_0}, \tag{65}$$

where Y is output and y_0 is the final good.

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

[3]We use normalized quantities since it simplifies calibration, and clarifies the fact that CES aggregators are not unit-less.

[4]For convenience we use number indices starting at 0 instead of 1 to describe the elements of ω and θ, but number indices starting at 1 to describe the elements of μ. We impose the restriction that $\omega_{ij} \in [0,1]$, $\sum_j \omega_{ij} = 1$ for all $0 \le i \le N$, $\omega_{fj} = 0$ for all $N < f \le N+F$, $\omega_{0f} = 0$ for all $N < f \le N+F$, and $\omega_{i0} = 0$ for all $0 \le i \le N$.

Consider some initial allocation with markups/wedges μ and productivity shifters normalized, without loss of generality, at $A = 1$. The normalizing constants $(\overline{y}, \overline{x})$ are chosen to correspond to this initial allocation. Let b and $\tilde{\Omega}$ be the corresponding vector of consumption shares and cost-based input-output matrix. Then we must have $\omega_{0i} = b_i$ and $\omega_{(i+1)(j+1)} = \tilde{\Omega}_{ij}$. From there, all the other cost-based and revenue-based input-output objects can be computed exactly as in Section 2.2.

Appendix H Robustness and Extensions

In this section, we discuss some of the extensions mentioned in the body of the paper. Specifically, we address in more detail how are results extend to situations with arbitrary non-CES production functions, elastic factors, capital accumulation/dynamics, and nonlinearities. Proofs for the results are at the end of this section.

H.1 Beyond CES

The input-output covariance operator defined in Section 4 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 j with cost function $\mathbf{C}_j$, we define the Allen-Uzawa elasticity of substitution between inputs x and y as

$$\theta_j(x, y) = \frac{\mathbf{C}_j d^2 \mathbf{C}_j / (dp_x dp_y)}{(d\mathbf{C}_j / dp_x)(d\mathbf{C}_j / dp_y)} = \frac{\epsilon_j(x, y)}{\Omega_{jy}},$$

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

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

We define the *input-output substitution operator* for producer j as

$$\Phi_j(\tilde{\Psi}_{(k)}, \Psi_{(f)}) = - \sum_{1 \le x, y \le N+1+F} \Omega_{jx} [\delta_{xy} + \Omega_{jy}(\theta_j(x, y) - 1)] \tilde{\Psi}_{xk} \Psi_{yf}, \tag{66}$$

$$= \frac{1}{2} E_{\Omega^{(j)}} \left((\theta_j(x, y) - 1)(\tilde{\Psi}_k(x) - \tilde{\Psi}_k(y))(\Psi_f(x) - \Psi_f(y)) \right), \tag{67}$$

where δ_{xy} is the Kronecker symbol, $\tilde{\Psi}_k(x) = \Psi_{xk}$ and $\Psi_f(x) = \Psi_{xf}$, 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_j, all the cross Allen-Uzawa elasticities are identical with $\theta_j(x, y) = \theta_j$ if $x \neq y$, and the own Allen-Uzawa elasticities are given by $\theta_j(x, x) = -\theta_j(1 - \Omega_{jx})/\Omega_{jx}$. It is easy to verify that we then recover the input-output covariance operator:

$$\Phi_j(\tilde{\Psi}_{(k)}, \Psi_{(f)}) = (\theta_k - 1)Cov_{\Omega^{(j)}}(\tilde{\Psi}_{(k)}, \Psi_{(f)}).$$

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: $\Phi_j(\tilde{\Psi}_{(k)}, \Psi_{(f)})$ is bilinear and symmetric in $\tilde{\Psi}_{(k)}$ and $\Psi_{(f)}$; $\Phi_j(\tilde{\Psi}_{(k)}, \Psi_{(f)}) = 0$ whenever $\tilde{\Psi}_{(k)}$ or $\Psi_{(f)}$ is a constant.

Luckily, it turns out that all of the results stated in Sections 4 and 5 can be generalized to non-CES economies simply by replacing terms of the form $(\theta_j - 1)Cov_{\Omega^{(j)}}(\tilde{\Psi}_{(k)}, \Psi_{(f)})$ by $\Phi_j(\tilde{\Psi}_{(k)}, \Psi_{(f)})$.

Intuitively, $\Phi_j(\tilde{\Psi}_{(k)}, \Psi_{(f)})$ captures the way in which j redirects demand expenditure towards f in response to proportional unit decline in the price of k. To see this, we make use of the following observation: the elasticity of the expenditure share of producer j on input x with respect to the price of input y is given by $\delta_{xy} + \Omega_{jy}(\theta_j(x, y) - 1)$. Equation (66) requires considering, for each pair of inputs x and y, how much the proportional reduction Ψ_{yk} in the price of y induced by a unit proportional reduction in the price of k causes producer j to increase its expenditure share on x (as measured by $-\Omega_{jx}[\delta_{xy} + \Omega_{jy}(\theta_j(x, y) - 1)]\tilde{\Psi}_{yk}$) and how much x is exposed to f (as measured by Ψ_{xf}).

Equation (67) says that this amounts to considering, for each pair of inputs x and y, whether or not increased exposure to k as measured by $\tilde{\Psi}_k(x) - \tilde{\Psi}_k(y)$, 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 $(\theta_j(x, y) - 1)$. If x and y are substitutes, and $\tilde{\Psi}_k(x) - \tilde{\Psi}_k(y)$ and $\Psi_f(x) - \Psi_f(y)$ are both positive, then substitution across x and y by k, in response to a shock to a decrease in the price of k, increases demand for f.

H.2 Elastic Factor Supplies

In this section, we fully flesh out one such extension by showing to generalize our analysis to allow for endogenous factor supplies.

To model elastic factor supplies, let $G_f(w_f Y, Y)$ be the aggregate supply of factor f, where $w_f Y$ is the price of the factor in the GDP deflator numeraire (w_f is the price of the factor in the nominal GDP numeraire) and Y is real aggregate income. Let $\zeta_f = \partial \log G_f / \partial \log(w_f Y)$ be the elasticity of the supply of factor f to its real wage, and $\gamma_f = -\partial \log G_f / \partial \log Y$ be its income elasticity. We then have the following characterization:

$$\frac{\mathrm{d}\log Y}{\mathrm{d}\log A_k} = \varrho\left(\tilde{\lambda}_k - \sum_f \frac{1}{1+\zeta_f}\tilde{\Lambda}_f \frac{\mathrm{d}\log \Lambda_f}{\mathrm{d}\log A_k}\right), \tag{68}$$

and

$$\frac{\mathrm{d}\log Y}{\mathrm{d}\log \mu_k} = \varrho\left(-\tilde{\lambda}_k - \sum_f \frac{1}{1+\zeta_f}\tilde{\Lambda}_f \frac{\mathrm{d}\log \Lambda_f}{\mathrm{d}\log \mu_k}\right), \tag{69}$$

where $\varrho = 1/(\sum_f \tilde{\Lambda}_f \frac{1+\gamma_f}{1+\zeta_f})$.

With inelastic factors, a decline in factor income shares, ceteris paribus, *increases* output since it represents a reduction in the misallocation of resources and an increase in aggregate TFP. With elastic factor supply, the output effect is dampened by the presence of $1/(1+\zeta_f) < 1$. This is due to the fact that a reduction in factor income shares, while increasing aggregate TFP, reduces factor payments and factor supplies, which in turn reduces output. Hence, when factors are elastic, increases in allocative efficiency from assigning more resources to more monopolistic producers are counteracted by reductions in factor supplies due to the associated suppression of factor demand.[5]

We can provide an explicit characterization of $\mathrm{d}\log \Lambda_f$ and $\mathrm{d}\log Y$ in terms of microeconomic elasticities of substitution in a nested-CES structure similar to the one in Section 4. Changes in factor shares and output solve the following system of equations:

$$\begin{aligned} d\log \Lambda_f = &-\sum_k \lambda_k \frac{\Psi_{kf}}{\Lambda_f} d\log \mu_k + \sum_j (\theta_j - 1)\frac{\lambda_j}{\mu_j} Cov_{\tilde{\Omega}^{(j)}}\Big(\sum_k \tilde{\Psi}_{(k)} d\log A_k - \sum_k \tilde{\Psi}_{(k)} d\log \mu_k, \frac{\Psi_{(f)}}{\Lambda_f}\Big) \\ &- \sum_j (\theta_j - 1)\frac{\lambda_j}{\mu_j} Cov_{\tilde{\Omega}^{(j)}}\Big(\sum_g \tilde{\Psi}_{(g)} \frac{1}{1+\zeta_g} d\log \Lambda_g + \sum_g \tilde{\Psi}_{(g)} \frac{\gamma_g - \zeta_g}{1+\zeta_g} d\log Y, \frac{\Psi_{(f)}}{\Lambda_f}\Big), \end{aligned}$$

$$d\log Y = \rho\left[\sum_k \tilde{\lambda}_k d\log A_k - \sum_k \tilde{\lambda}_k d\log \mu_k - \sum_f \tilde{\Lambda}_f \frac{1}{1+\zeta_f} d\log \Lambda_f\right].$$

[5] In the limit where factor supplies become infinitely elastic, the influence of the allocative efficiency effects disappear from output, since more factors can always be marshaled on the margin at the same real price. To see this, consider the case with a single factor called labor, and factor supply function $G_L(wY, Y) = Y^{-\nu}(wY)^{\nu}$, which can be derived from a standard labor-leisure choice model. In this case, $\zeta_L = \gamma_L = \nu$, and so equation (68) implies that $\mathrm{d}\log Y/\mathrm{d}\log A_k = \tilde{\lambda}_k - 1/(1+\nu)\,\mathrm{d}\log \Lambda_L/\mathrm{d}\log A_k$. When labor supply becomes infinitely elastic $\nu \to \infty$, this simplifies to $\mathrm{d}\log Y/\mathrm{d}\log A_k = \tilde{\lambda}_k$, so that changes in allocative efficiency have no effect on output, even though they affect TFP.

Equations (68) and (69) can also be applied to frictionless economies with endogenous factor supplies. They show that even without any frictions, Hulten's theorem cannot be used to predict how output will respond to microeconomic TFP shocks, due to endogenous responses of factors. These results therefore also extend Hulten's theorem to efficient economies with endogenous factor supplies.

H.3 Capital Accumulation, Adjustment Costs, and Capacity Utilization

In mapping this set-up to the data, there are two ways to interpret this model: either we could interpret final demand as a per-period part of a larger dynamic problem, or we could interpret final demand as an intertemporal consumption function where goods are also indexed by time à la Arrow-Debreu. When we interpret the model intratemporally, the output function encompasses demand for consumption goods and for investment goods. When we interpret the model intertemporally, the process of capital accumulation is captured via intertemporal production functions that transform goods in one period into goods in other periods. This modeling choice would also be well-suited to handle technological frictions to the reallocation of factors such as adjustment costs and variable capacity utilization. Our formulas would apply to these economies without change, but of course, in such a world, input-output definitions would now be expressed in net-present value terms.

H.4 Nonlinear Impact of Shocks and Duality with Industry Structure

Another limitation of our results is that we neglect nonlinearities. As discussed by Baqaee and Farhi (2017a), models with production networks can respond very nonlinearly to productivity shocks. We plan to extend these results to inefficient economies in full generality, but as a first step, we stipulate some conditions under which we can directly leverage these results to inefficient economies. In particular, we show that the amplification of negative shocks due to complementarities emphasized in Baqaee and Farhi (2017a) can also work to amplify the negative effects of misallocation.

Consider the quantitative parametric model in Section 7. Let $\delta_k(i)$, $\mu_k(i)$, and $A_k(i)$ denote firm i in industry k's share of industry sales, markup, and productivity. Define industry k's average markup and productivity to be

$$\mu_k = \left(\sum_i \frac{\delta_k(i)}{\mu_k(i)} \right)^{-1}$$

and

$$A_k/\overline{A_k} = \frac{\mu_k/\overline{\mu_k}}{\left(\sum_i \overline{\delta_k(i)} \left(\frac{\mu_k(i)/\overline{\mu_k(i)}}{A_k(i)/\overline{A_k(i)}}\right)^{1-\xi_k}\right)^{\frac{1}{1-\xi_k}}},$$

where overline variables denote steady-state values.

To the original firm-level economy, we associate a dual industry-level economy, for which the input output-matrix is aggregated at the industry level. Define the output level of the dual economy by $\check{Y}$ and the revenue-based Domar weight of industry k by $\check{\lambda}_k$. The dual industry-level economy has initial industry-level markups equal to

$$\overline{\mu}_k = \left(\sum_i \frac{\overline{\delta}_k(i)}{\mu_k(i)}\right)^{-1}.$$

The next proposition shows that productivity and markup shocks in the original firm-level economy can be translated into productivity and markup shocks in the dual industry-level economy.

Proposition 6 (Exact Duality)**.** *The discrete (nonlinear) output response* $\Delta \log Y$ *to shocks to productivities and markups of the original economy is equal to the output response* $\Delta \log \check{Y}$ *to the dual shocks to productivity and markups of the dual economy.*

Corollary 3 (Efficient Duality)**.** *Consider an economy where* $\overline{\mu_k} = 1$ *for every* k*, and consider a transformation* $\mu_k(i)(t_k)$ *which changes markups but maintains* $\mu_k = 1$*. Then*

$$\frac{d\log Y}{d\log t_k} = \frac{d\log \check{Y}}{d\log t_k} = \check{\lambda}_k \frac{d\log \check{A}_k}{d\log t_k} \tag{70}$$

and

$$\frac{d^2\log Y}{d\log t_k^2} = \frac{d^2\log \check{Y}}{d\log t_k^2} = \check{\lambda}_k \frac{d\log \check{\lambda}_k}{d\log \check{A}_k}\left(\frac{d\log \check{A}_k}{d\log t_k}\right)^2 + \check{\lambda}_k \frac{d^2\log \check{A}_k}{d\log t_k^2}, \tag{71}$$

where $d\log \check{\lambda}_k / d\log \check{A}_k$ *is given by the formulas in Baqaee and Farhi (2017a).*

If firms within an industry are substitutes, then increases in the dispersion of markups, which keep the harmonic average of markups equal to one, are isomorphic to negative productivity shocks in a model which is efficient at the industry level. Hence, shocks which increase markup dispersion in an industry can have outsized nonlinear effects on output, if those industries are macro-complementary with other industries in the sense defined by Baqaee and Farhi (2017a) so that $d\log \check{\lambda}_k / d\log \check{A}_k < 0$.

This helps flesh out the insight in Jones (2011) that complementarities can interact with distortions to generate large reductions in output, and that these can be quantitatively important enough to explain the large differences in cross-country incomes. Given the examples in Baqaee and Farhi (2017a), it should be clear how misallocation in a key industry like energy production can significantly reduce output through macro-complementarities. Investigating these nonlinear forces more systematically is an interesting exercise that we leave for future work.

Proof of Proposition 6. To streamline the exposition, we focus on a single industry, and we use different but more straightforward notation. Consider an industry where: all firms i use the same upstream input bundle with cost C; firms transform this input into a firm-specific variety of output using constant return to returns to scale technology; each firm i has productivity a_i and charges a markup μ_i; the varieties are combined into a composite good by a competitive downstream industry according to a CES production function with elasticity σ on firm i. Without loss of generality, and only for convenience, we normalize all prices in steady-state to be equal to C, which means that we normalize the levels of productivities in steady state (and only in steady state) to be equal to the markups.

We denote the quantity of composite good produced as

$$Q = \left[\sum b_i^{\frac{1}{\sigma}} q_i^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}. \tag{72}$$

Firm i charges a price

$$p_i = \frac{\mu_i}{a_i} C. \tag{73}$$

The resulting demand for firm i's variety is

$$q_i = (\frac{p_i}{P})^{-\sigma} b_i Q, \tag{74}$$

where the price index is given by

$$P = \left[\sum b_i p_i^{1-\sigma}\right]^{\frac{1}{1-\sigma}}. \tag{75}$$

Total profits are given by

$$\Pi = \sum_i (p_i - C)(\frac{p_i}{P})^{-\sigma} b_i Q. \tag{76}$$

We solve out the price index and profits explicitly and get

$$P = \left[\sum_i b_i \left(\frac{\mu_i}{a_i} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} C, \tag{77}$$

$$\Pi = \sum_i \left(\frac{\mu_i}{a_i} - \frac{1}{a_i} \right) \left[\frac{\frac{\mu_i}{a_i}}{\left[\sum_j b_j \left(\frac{\mu_j}{a_j} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}} \right]^{-\sigma} b_i CQ. \tag{78}$$

For completeness we can also solve for the sales of each firm as a fraction of the sales of the industry

$$\lambda_i = \frac{p_i q_i}{PQ} = \frac{b_i \left(\frac{\mu_i}{a_i} \right)^{1-\sigma}}{\sum_j b_j \left(\frac{\mu_j}{a_j} \right)^{1-\sigma}}. \tag{79}$$

We want to understand how to aggregate this industry into homogenous industry with productivity A and markup μ. These variables must satisfy

$$P = \frac{\mu}{A} C, \tag{80}$$

$$\Pi = \left(\frac{\mu}{A} - \frac{1}{A} \right) CQ. \tag{81}$$

This implies that A and μ are the solutions of the following system of equations

$$\frac{\mu}{A} = \left[\sum_i b_i \left(\frac{\mu_i}{a_i} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}, \tag{82}$$

$$\left(\frac{\mu}{A} - \frac{1}{A} \right) = \sum_i \left(\frac{\mu_i}{a_i} - \frac{1}{a_i} \right) \left[\frac{\frac{\mu_i}{a_i}}{\left[\sum_j b_j \left(\frac{\mu_j}{a_j} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}} \right]^{-\sigma} b_i. \tag{83}$$

The solution is

$$A = \frac{1}{\left[\sum_i b_i \left(\frac{\mu_i}{a_i} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} - \sum_i \left(1 - \frac{1}{\mu_i} \right) \left[\frac{\frac{\mu_i}{a_i}}{\left[\sum_j b_j \left(\frac{\mu_j}{a_j} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}} \right]^{-\sigma} \frac{\mu_i}{a_i} b_i}, \tag{84}$$

$$\mu = \frac{\left[\sum_i b_i \left(\frac{\mu_i}{a_i}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}}{\left[\sum_i b_i \left(\frac{\mu_i}{a_i}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}} - \sum_i \left(1 - \frac{1}{\mu_i}\right)\left[\frac{\frac{\mu_i}{a_i}}{\left[\sum_j b_j \left(\frac{\mu_j}{a_j}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}}\right]^{-\sigma} \frac{\mu_i}{a_i} b_i}. \tag{85}$$

We can also rewrite this in a useful way as

$$A = \frac{1}{\left[\sum_i b_i \left(\frac{\mu_i}{a_i}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}} \frac{1}{\sum_i \frac{1}{\mu_i} \frac{\left(\frac{\mu_i}{a_i}\right)^{1-\sigma} b_i}{\sum_j b_j \left(\frac{\mu_j}{a_j}\right)^{1-\sigma}}} = \frac{1}{\left[\sum_i b_i \left(\frac{\mu_i}{a_i}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}} \frac{1}{\sum_i \frac{1}{\mu_i} \lambda_i}, \tag{86}$$

$$\mu = \frac{1}{\sum_i \frac{1}{\mu_i} \frac{\left(\frac{\mu_i}{a_i}\right)^{1-\sigma} b_i}{\sum_j b_j \left(\frac{\mu_j}{a_j}\right)^{1-\sigma}}} = \frac{1}{\sum_i \frac{1}{\mu_i} \lambda_i}. \tag{87}$$

The theorem follows by applying this analysis to an industry k, with $\delta_k(i) = \lambda_i$ and $b_i = \overline{\delta_k(i)}$.

■

Appendix I Aggregation of Cost-Based Domar Weights

In this Appendix we show that recovering cost-based Domar weights from aggregated data is, in principle, not possible. The vertical economy in Figure 1a also shows the failure of the aggregation property implied by Hulten's theorem. The easiest way to see this is to consider aggregating the input-output table for the economy in Figure 1a. For simplicity, suppose that markups are the same everywhere so that $\mu_i = \mu$ for all i. Since there is no possibility of reallocation in this economy, and since markups are uniform, this is our best chance of deriving an aggregation result, but even in this simplest example, such a result does not exist. Suppose that we aggregate the whole economy $S = \{1, \ldots, N\}$. Then, in aggregate, the economy consists of a single industry that uses labor and inputs from itself to produce. In this case, the input-output matrix is a scalar, and equal to the intermediate input share of the economy

$$\Omega_{SS} = \frac{1 - \frac{1}{\mu^{N-1}}}{1 - \frac{1}{\mu^N}}, \tag{88}$$

and the aggregate markup for the economy is given by μ. Therefore, $\tilde{\lambda}_S$ constructed using aggregate data is

$$\tilde{\lambda}_S = 1'(I - \mu\Omega)^{-1} = \frac{\mu^{N-1} - \frac{1}{\mu}}{1 - \frac{1}{\mu}}. \tag{89}$$

However, we know from the example that

$$\frac{\mathrm{d}\log Y}{\mathrm{d}\log A} = \sum_{i \in S} \tilde{\lambda}_i = N \neq \tilde{\lambda}_S = \frac{\mu^{N-1} - \frac{1}{\mu}}{1 - \frac{1}{\mu}}, \tag{90}$$

except in the limiting case without distortions $\mu \to 1$. Therefore, even in this simplest case, with homogenous markups and no reallocation, aggregated input-output data cannot be used to compute the impact of an aggregated shock.

Appendix J Extra Examples

Example J.1. We build a simple example to underscore the importance of properly accounting for the multiplicity of factors to assess the macroeconomic impact of microeconomic shocks in inefficient economies. The example is depicted in Figure 10.

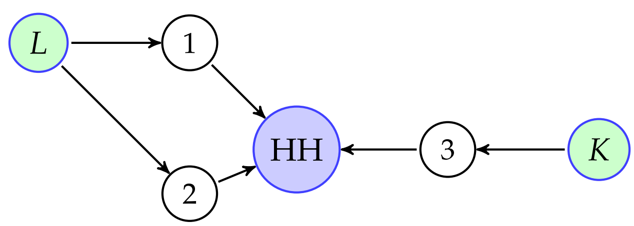

Figure 10: An economy with two factors of production L and K. The subgraph from L to the household contains a cycle, and hence can be subject to misallocation. On the other hand, there is only a unique path connecting K to the household, so there is no misallocation.

We have

$$\Gamma = -(\theta_0 - 1)\begin{pmatrix} Cov_b(\tilde{\Psi}_{(L)}, \Psi_{(L)}) & Cov_b(\tilde{\Psi}_{(K)}, \Psi_{(L)}) \\ Cov_b(\tilde{\Psi}_{(L)}, \Psi_{(K)}) & Cov_b(\tilde{\Psi}_{(K)}, \Psi_{(K)}) \end{pmatrix}, \tag{91}$$

and

$$\delta_{(i)} = (\theta_0 - 1)\begin{pmatrix} Cov_b(\tilde{\Psi}_{(i)}, \Psi_{(L)}) \\ Cov_b(\tilde{\Psi}_{(i)}, \Psi_{(K)}) \end{pmatrix}.$$

Substituting in the values and solving the system of equations (11), using Proposition 2, and noting that $\lambda_i = \tilde{\lambda}_i$ for all i, we find that

$$\frac{\mathrm{d}\log Y}{\mathrm{d}\log A_i} = \lambda_i + \lambda_i(\theta_0 - 1)\left(1 - \frac{\mu_i^{-1}}{\frac{\lambda_1}{\lambda_1+\lambda_2}\mu_1^{-1} + \frac{\lambda_2}{\lambda_1+\lambda_2}\mu_2^{-1}}\right), \quad (i = 1, 2)$$

but

$$\frac{\mathrm{d}\log Y}{\mathrm{d}\log A_i} = \lambda_i, \quad (i = 3).$$

A lesson is that changes in allocative efficiency are only present for shocks to producers 1 and 2 which share a factor of production, but not for producer 3 which has its own factor of production. Moreover, the changes in allocative efficiency for shocks to producers 1 and 2 only depends on the markups of these two producers and not on the markup of producer 3.

Example J.2. We consider a simple example with two elasticities of substitution, which demonstrates the principle that changes in misallocation are driven by how each node switches its demand across its supply chain in response to a shock. To this end, we apply Proposition 2 to the economy depicted in Figure 11.

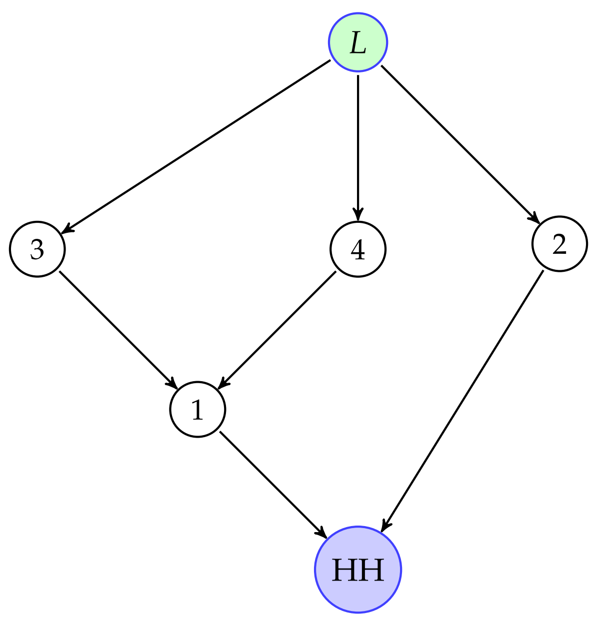

Figure 11: An economy with two elasticities of substitution.

$$\begin{aligned}\frac{\mathrm{d}\log Y}{\mathrm{d}\log A_3} =& \tilde{\lambda}_3 - \frac{1}{\Lambda_L}\Big((\theta_0 - 1)\left[b_1(\omega_{13}\mu_1^{-1}(\omega_{13}\mu_3^{-1} + \omega_{14}\mu_4^{-1})) - \omega_{13}b_1\Lambda_L\right] \\ & +(\theta_1 - 1)\mu_1^{-1}\lambda_1\left[\omega_{13}\mu_3^{-1} - \omega_{13}\left(\omega_{13}\mu_3^{-1} + \omega_{14}\mu_4^{-1}\right)\right]\Big).\end{aligned}$$

The term multiplying $(\theta_0 - 1)$ captures how the household will shift their demand across 1 and 2 in response to the productivity shock, and the relative degrees of misallocation in 1 and 2's supply chains. The term multiplying $(\theta_1 - 1)$ takes into account how 1 will shift its demand across 3 and 4 and the relative amount of misallocation of labor between 3 and 4. Not surprisingly, if instead we shock industry 1, then only the household's elasticity of substitution matters, since industry 1 will not shift its demand across its inputs in response to the shock to industry 2:

$$\frac{\mathrm{d}\log Y}{\mathrm{d}\log A_1} = b_1 - \frac{1}{\Lambda_L}(\theta_0 - 1)\left[b_1\mu_1^{-1}(\omega_{13}\mu_3^{-1} + \omega_{14}\mu_4^{-1}) - b_1\Lambda_L\right].$$

This illustrates the general principle in Proposition 2 that an elasticity of substitution θ_j matters only if j is somewhere downstream from k.

Example J.3. We build a simple example to illustrate the macroeconomic impact of microeconomic markup/wedge shocks and their difference with microeconomic productivity shocks.

We consider a Cobb-Douglas economy, which helps to isolate the importance of the new term in Proposition 3. For a Cobb-Douglas economy, the only source factor reallocation comes from the fact that the producer which increases its markup/wedge releases some labor. Let $\theta_j = 1$ for every j, which is the Cobb-Douglas special case. Now, applying Proposition 3, we get

$$\frac{\mathrm{d}\log Y}{\mathrm{d}\log \mu_k} = -\tilde{\lambda}_k + \lambda_k\frac{\Psi_{kL}}{\Lambda_L} = -\tilde{\lambda}_k\left(1 - \frac{\lambda_k}{\tilde{\lambda}_k}\frac{\Psi_{kL}}{\Lambda_L}\right).$$

As before, Ψ_{kL}/Λ_L is a measure of how distorted the supply chain of k is relative to the economy as a whole. If $\Psi_{kL}/\Lambda_L < 1$, then this means that for each dollar k earns, a smaller share reaches workers than it would if that dollar was spent by the household. In other words, producer k's supply chain has inefficiently too few workers. On the other hand, $\lambda_k/\tilde{\lambda}_k$ is a measure of how distorted the demand of chain of k is. If $\lambda_k/\tilde{\lambda}_k < 1$, this implies that k is facing double-marginalization. When the product of the downstream and upstream terms is less than one, this means that producer k is inefficiently starved of demand and workers. Hence, an increase in the markup/wedge of k reduces the allocative efficiency of the economy. On the other hand, when the product of these two terms is greater than one, the path connecting the household to labor via producer k is too large. Therefore, an increase in the markup/wedge of k reallocates resources to the rest of the economy where they are more needed and increases allocative efficiency.

With multiple factors, we get

$$\frac{\mathrm{d}\log Y}{\mathrm{d}\log \mu_k} = -\tilde{\lambda}_k + \lambda_k \sum_f \tilde{\Lambda}_f \frac{\Psi_{kf}}{\Lambda_f} = -\tilde{\lambda}_k \left(1 - \frac{\lambda_k}{\tilde{\lambda}_k} \sum_f \tilde{\Lambda}_f \frac{\Psi_{kf}}{\Lambda_f} \right).$$

This generalizes the intuitions discussed earlier for markup/wedge shocks in the Cobb-Douglas economy with a single factor to the case of multiple factors. In particular, the amount of factor f released by sector k as a fraction of total factor f per unit of shock is $\lambda_k \Psi_{kf}/\Lambda_f$ and the impact of that release on output per unit of shock is $\tilde{\Lambda}_f$. We also see again the roles of the index of downstream distortions $\lambda_k/\tilde{\lambda}_k$ and of the generalized index of upstream distortions $\sum_f \tilde{\Lambda}_f \Psi_{kf}/\Lambda_f$.

Example J.4. We consider an example showing how, in general, the correlation between productivity and wedges matters. Consider the horizontal economy example discussed in Section 5.1, but instead of assuming log-normality, consider the binomial case where $A_i \in \{0, A\}$ with probability 1/2 and $\Delta \log \mu_i \in \{0, \Delta \log \mu\}$ with probability 1/2. An immediate application of formula (19) shows that $\mathcal{L} \approx (1/8)\theta_0(\Delta \log \mu)^2$ if A_i and μ_i are independent, but that $\mathcal{L} \approx 0$ if A_i and μ_i are perfectly correlated.

Appendix K Volatility of Aggregate TFP

In this section, we use the quantitative structural model of Section 7.3 to assess the volatility of aggregate output arising from firm-level and industry-level productivity and markup shocks.[6] For this section, we do not assume that each Compustat firms' share of industry sales in Compustat is the same as its share of total industry sales in the BEA data. Instead, we assign to each firm its actual sales, and assume that any leftover sales are sold by a residual producer whose markup is equal to the average industry-level markup and who experiences no shocks (this effectively means we assume that the residual (non-Compustat) producer in each industry is really a representative of a mass of infinitesimal firms and experiences no shocks due to the law of large numbers).

We use our ex-post structural results on the elasticities of aggregate output to these shocks

$$\log Y \approx \log \overline{Y} + \sum_i \frac{\mathrm{d}\log Y}{\mathrm{d}\log A_i} \mathrm{d}\log A_i + \sum_i \frac{\mathrm{d}\log Y}{\mathrm{d}\log \mu_i} \mathrm{d}\log \mu_i,$$

[6] When we consider firm-level shocks, we assess only the contribution of shocks to Compustat firms, i.e. we account for the macro-volatility arising from firm-level shocks when only Compustat firms are being shocked, and not non-Compustat firms. We focus on this exercise because we do not have the data required to compute the contribution of shocks to all firms.

to approximate the implied volatility of output in response to microeconomic shocks. Assuming productivity shocks and markup shocks are independent and identically distributed, get

$$
\begin{aligned}
Var(\log Y) &\approx \sum_i \left(\frac{\mathrm{d}\log Y}{\mathrm{d}\log A_i}\right)^2 Var(\mathrm{d}\log A_i) + \sum_i \left(\frac{\mathrm{d}\log Y}{\mathrm{d}\log \mu_i}\right)^2 Var(\mathrm{d}\log \mu_i), \\
&= \|D_{\log A}\log Y\|^2 Var(\mathrm{d}\log A) + \|D_{\log \mu}\log Y\|^2 Var(\mathrm{d}\log \mu).
\end{aligned}
$$

Hence, the Euclidean norm $\|D_{\log A}\log Y\|$ of the Jacobian of $\log Y$ with respect to $\log A$ gives the degree to which microeconomic productivity shocks are not "diversified" away in the aggregate. Similarly, $\|D_{\log \mu}\log Y\|$ measures the diversification factor relative to markup shocks.[7]

Table 3 displays the diversification factor, for both markup shocks and productivity shocks at the firm level and at the industry level, for our benchmark model. We also compute the results for a Cobb-Douglas distorted economy where all elasticities are unitary, as well as for a perfectly competitive model without wedges. Across the board, the distorted model is more volatile than the competitive model, however the extent of this depends greatly on the type of shock and the level of aggregation. We discuss these different cases in turn.

First, consider the case of productivity shocks: as mentioned previously, the benchmark model is more volatile than the perfectly competitive model for both sets of shocks. However, the more interesting comparison is with respect to the distorted Cobb-Douglas economy. As explained in Section 4, the allocation of factors is invariant to productivity shocks in the Cobb-Douglas model. Hence, the Cobb-Douglas model lacks the reallocation channel, and hence can tell us in which direction the reallocation force is pushing. In the case of industry-level shocks, the benchmark model is slightly less volatile than the Cobb-Douglas model, whereas in the case of firm-level shocks, the benchmark model is significantly more volatile.

A partial intuition here relates to the elasticities of substitution: whereas industries are complements, firms within an industry are strong substitutes. Recall that loosely speaking, changes in allocative efficiency scale with the elasticity of substitution minus one. Firm-level shocks cause a considerable amount of changes in allocative efficiency whereas industry-level shocks cause much milder changes. At both levels of aggregation,

[7] Although Baqaee and Farhi (2017a) suggest that log-linear approximations can be unreliable for modeling the mean, skewness, or kurtosis of output in the presence of microeconomic shocks, their results indicate the log-linear approximations of variance are less fragile (although still imperfect). In the final section of this paper, we discuss how our results can be extended to understanding the nonlinear impact of shocks.

	Benchmark	Competitive	Cobb-Douglas	Constant $\mathcal{X}$
Firm Productivity Shocks (UC)	0.0491	0.0376	0.0396	0.0396
Firm Markup Shocks (UC)	0.0296	0.0000	0.0077	0.0000
Industry Productivity Shocks (UC)	0.3162	0.3118	0.3259	0.3259
Industry Markup Shocks (UC)	0.0084	0.0000	0.0391	0.0000
Firm Productivity Shocks (AP)	0.0524	0.0376	0.0415	0.0415
Firm Markup Shocks (AP)	0.0368	0.0000	0.0085	0.0000
Industry Productivity Shocks (AP)	0.3188	0.3118	0.3375	0.3375
Industry Markup Shocks (AP)	0.0127	0.0000	0.0500	0.0000
Firm Productivity Shocks (PF)	0.0598	0.0376	0.0398	0.0398
Firm Markup Shocks (PF)	0.0321	0.0000	0.0112	0.0000
Industry Productivity Shocks (PF)	0.3299	0.3418	0.3618	0.3618
Industry Markup Shocks (PF)	0.0216	0.0000	0.0760	0.0000

Table 3: Diversification factor for different productivity and markup shocks at firm and industry level for different specifications of the model. A diversification factor of 1 means that the variance of microeconomic shocks moves aggregate variance one-for-one. A diversification factor of 0 means that microeconomic shocks are completely diversified away at the aggregate level. The final column is the allocation that holds the allocation matrix $\mathcal{X}$ constant in response to shocks.

these changes in allocative efficiency amplify some shocks and mitigate some others compared to the Cobb-Douglas model with no change in allocative efficiency.[8] On the whole, at the firm level, the changes in allocative efficiency are so large that they dwarf the pure technology effects picked up by the Cobb-Douglas model and amplify the volatility of these shocks. By contrast, at the industry level, changes in allocative efficiency are more moderate and turn out to slightly mitigate the volatility of these shocks.

These intuitions are confirmed in the first two columns of Figure 12, where we plot the output elasticity with respect to productivity shocks to specific firms or industries relative to their revenue-based Domar weight. This represents a comparison of our benchmark model to a competitive model where Hulten's theorem holds. We find considerable dispersion in the response of the model relative to both, but much more so at the firm level than at the industry level. We could plot the same graph but with the cost-based Domar weight as a reference point in order to represent the comparison of our benchmark model to the Cobb-Douglas model, and the results would be visually similar.

Next, consider the effects of markup shocks. In this case, the distorted Cobb-Douglas economy is not necessarily a very natural benchmark since even with Cobb-Douglas, shocks to markups will reallocate factors across producers. Nonetheless, it is still instructive to compare the benchmark model to the Cobb-Douglas one to find that a similar lesson applies as with productivity shocks. The volatility of firm-level shocks is amplified relative to Cobb-Douglas while the volatility of industry-level shocks is attenuated relative to Cobb-Douglas. This follows from the fact that industries are more complementary than firms, and hence, in line with the intuition from the horizontal economy, the effect of the shock are monotonically increasing in the degree of substitutability. The last two columns of Figure 12 plot the output elasticity with respect to markup shocks to specific firms or industries relative to their revenue-based Domar weight.

Online References

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

Atkeson, A. and A. Burstein (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review 98*(5), 1998–2031.

[8]There is another difference: reallocation occurs towards the firm receiving a positive shock; but reallocation occurs away from the industry receiving a positive shock.

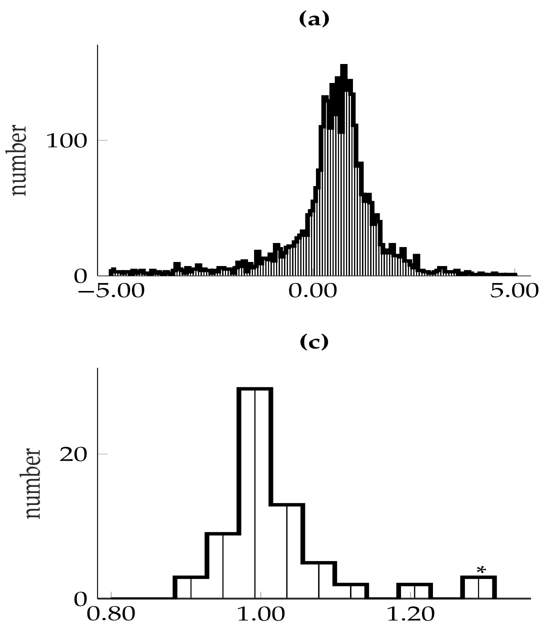

Figure 12: The left column contains histograms of $\mathrm{d}\log Y/\,\mathrm{d}\log A$ and the right $\mathrm{d}\log Y/\,\mathrm{d}\log \mu$ relative to λ for firm-level and industry-level shocks respectively. The top row are shocks to firms and the bottom row are shocks to industries. The bunching at the extremes, winsorizing at 4 standard deviations, marked with a star arise solely for displaying purposes. In all cases, the degree of dispersion around the response implied by the competitive model (the size of the producer) is substantial.

Baqaee, D. R. and E. Farhi (2017). The macroeconomic impact of microeconomic shocks: Beyond Hulten's Theorem.

Claus, J. and J. Thomas (2001). Equity premia as low as three percent? evidence from analysts' earnings forecasts for domestic and international stock markets. *The Journal of Finance 56*(5), 1629–1666.

De Loecker, J., J. Eeckhout, and G. Unger (2019). The rise of market power and the macroeconomic implications. Technical report, National Bureau of Economic Research.

Gutiérrez, G. and T. Philippon (2016). Investment-less growth: An empirical investigation. Technical report, National Bureau of Economic Research.

Hall, R. E. (1988). The relation between price and marginal cost in us industry. *Journal of political Economy 96*(5), 921–947.

Jones, C. I. (2011). Intermediate goods and weak links in the theory of economic development. *American Economic Journal: Macroeconomics*, 1–28.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The review of economic studies 70*(2), 317–341.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica: Journal of the Econometric Society*, 1263–1297.