Online Appendix to
Entry vs. Rents

David Rezza Baqaee[*] Emmanuel Farhi[†]

A Endogenous Markups **48**

B Proofs **49**

C The Role of Reallocation **63**

D Beyond CES **65**

E Details of Sectoral Models **66**

F Relaxing Homotheticity/Iso-elasticity **72**

G Mapping Model to Data **80**

H Additional Quantitative Results **85**

I Second-Best Policy **85**

J Comparison to Simplified Models **87**

---

[*]Email: baqaee@econ.ucla.edu

[†]Emmanuel Farhi tragically passed away in July, 2020. He was a once-in-a-lifetime friend and collaborator. We thank Pol Antras, Natalie Bau, Ariel Burstein, Swati Dhingra, Elhanan Helpman, Hugo Hopenhayn, Oleg Itskhoki, Xavier Gabaix, Marc Melitz, Alireza Tahbaz-Salehi, Felix Tintelnot, and especially Jonathan Vogel for comments. Veronica De Falco provided exceptional research assistance. We acknowledge research financial support from the Ferrante fund at Harvard University and NSF grant #1947611.

Appendix A Endogenous Markups

In this section, we show how our results can be extended to economies with endogenous monopolistically-competitive markups. Recall that for each $i \in \mathcal{N}^{IRS}$, the consumer-surplus ratio is given by

$$\gamma_i - 1 = \frac{F_i(\frac{y_i}{Y_i})}{F_i'(\frac{y_i}{Y_i})\frac{y_i}{Y_i}} \geq 0.$$

The price elasticity of demand is given by

$$\sigma_i = -\frac{F_i'(\frac{y_i}{Y_i})}{F_i''(\frac{y_i}{Y_i})\frac{y_i}{Y_i}},$$

where $\sigma_i > 1$ if marginal revenue is positive. If marginal revenue is strictly decreasing ($xF_i'''(x) < -2F_i''(x)$ for every x), then the monopolistically-competitive markup is uniquely determined by

$$\mu_i = \frac{1}{1 - \frac{1}{\sigma_i}} \geq 1.$$

The pass-through of marginal cost into the price is given by

$$\rho_i = 1 + \frac{\partial \log \mu_i}{\partial \log mc_i} = \frac{1}{1 - \mu_i \frac{\frac{y}{Y}\sigma_i'(\frac{y}{Y})}{\sigma_i(\frac{y}{Y})}} > 0.$$

Pass-through is greater than zero as long as marginal revenue curves are strictly downward sloping.

Lemma 3. *In the monopolistically-competitive equilibrium, the change in the markup of each* $i \in \mathcal{N}^{IRS}$ *is given by*

$$d\log \mu_i = -\frac{1-\rho_i}{\rho_i}\frac{\gamma_i}{\sigma_i} d\log M_i.$$

Hence, as long as pass-through is incomplete ($\rho_i < 1$), then an increase in the mass of firms of type i will cause markups in i to decline. Proposition 5 uses Lemma 3 to provide a version of Theorem 3 with endogenous markups.

Proposition 5 (Output Response with Inefficiencies and Endogenous Markups)**.** *Assume monopolistically competitive markups in every* $i \in \mathcal{N}$*. The response of aggregate output to shocks* $\mathrm{d}\log A$ *is given by*

$$\mathrm{d}\log Y = \sum_i \lambda_i^F d\log A_i - \sum_{i \in \mathcal{N}^{IRS}} \lambda_i^F \frac{1-\rho_i}{\rho_i}\frac{\gamma_i}{\sigma_i}\widehat{d\log \lambda}_{\pi,i} \tag{22}$$

$$- \sum_{i \in \mathcal{N}^{DRS}} \lambda_i^F (1 - \varepsilon_i) \left(d \log \lambda_{\pi,i} - \widehat{d \log \lambda}_{\pi,i} \right) + \sum_{i \in \mathcal{N}^{IRS}} \lambda_i^F (\gamma_i - 1) \widehat{d \log \lambda}_{\pi,i},$$

where we redefine Ψ^F to be

$$\Psi^F = \left(I - \mu \Omega^V - \left[(\gamma - \varepsilon) + \frac{1-\rho}{\rho} \frac{\gamma}{\sigma} \varepsilon \right] \tilde{\zeta}' \left(\tilde{\zeta} \lambda_\pi \tilde{\zeta}' \right)^{-1} \lambda_E \Omega^E \right)^{-1}.$$

There are two differences between Theorem 3, where markups are exogenous, and Proposition 5. First, the definition of the forward Leontief inverse has been modified; second, the term $d \log \mu_i$ has been replaced by $\frac{1-\rho_i}{\rho_i} \frac{\gamma_i}{\sigma_i} \widehat{d \log \lambda}_{\pi,i}$. We discuss each in turn.

The modification of Ψ^F accounts for the fact that a change in the price of j can affect the price of i by affecting the costs of entry into i via changes in markups. In particular, an increase in the price of j can raise the entry costs for entering into i, this reduces the mass of firms in i. If pass-through is incomplete, $\rho_i < 1$, then this raises the price of i and the new definition of Ψ^F accounts for this fact.

The appearance of $\frac{1-\rho_i}{\rho_i} \frac{\gamma_i}{\sigma_i} \widehat{d \log \lambda}_{\pi,i}$ in place of $d \log \mu_i$ captures the fact that an increase in profitability of i, if it raises quasi-rents, will cause markups in i to change endogenously. If pass-through is incomplete $\rho_i < 1$, then this causes markups to fall.

Forward propagation, Proposition 1, can likewise be modified.

Proposition 6 (Forward Propagation with Endogenous Markups)**.** *Assume monopolistically competitive markups in every $i \in \mathcal{N}$. In response to shocks* $\mathrm{d} \log A$*, changes in prices are given by*

$$\begin{aligned} d \log P_i = &- \sum_{j \in \mathcal{N}} \Psi_{ij}^F d \log A_j + \sum_{j \in \mathcal{N}^{IRS}} \Psi_{ij}^F \frac{1-\rho_i}{\rho_i} \frac{\gamma_i}{\sigma_i} \widehat{d \log \lambda}_{\pi,i} \\ &+ \sum_{j \in \mathcal{N}^{DRS}} \Psi_{ij}^F \left(1 - \varepsilon_j \right) \left(d \log \lambda_{\pi,j} - \widehat{\mathrm{d} \log \lambda}_{\pi,j} \right) - \sum_{j \in \mathcal{N}^{IRS}} \Psi_{ij}^F \left(\gamma_j - 1 \right) \widehat{\mathrm{d} \log \lambda}_{\pi,j}, \end{aligned}$$

where Ψ^F is defined according to Proposition 5.

Backward propagation, Proposition 2, is unchanged.

Appendix B Proofs

Proof of Theorem 1. Consider the Kuhn-Tucker conditions for the social planning problem

$$\mathcal{L} = U(C_1, \ldots, C_N) + \sum_{i \in N} \rho_i \left(Y_i - C_i - \sum_{k \in N} M_k x_{ki} - \sum_{j \in E} x_{E,ji} \right) + \sum_{i \in N} \delta_i \left(M_i F_i \left(\frac{y_i}{Y_i} \right) - 1 \right)$$

$$+\sum_{i\in N}\alpha_i\left(A_if_i(x_{ij}-y_i\right)+\sum_{j\in E}\beta_j\left(g_j\left\{x_{E,ji}\right\}-M_{E,j}\right)+\sum_{i\in N}\kappa_i\left(\sum_{j\in E}\zeta(j,i)M_{E,j}-M_i\right).$$

The first order conditions are

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial C_i} &: \frac{\partial U}{\partial C_i}-\rho_i=0,\\
\frac{\partial\mathcal{L}}{\partial Y_i} &: \rho_i-\delta_iM_iF_i'(\frac{y_i}{Y_i})\frac{y_i}{Y_i}\frac{1}{Y_i}=0,\\
\frac{\partial\mathcal{L}}{\partial x_{ij}} &: -\rho_jM_i+\alpha_iA_i\frac{\partial f_i}{\partial x_{ij}}=0,\\
\frac{\partial\mathcal{L}}{\partial M_i} &: -\sum_{k\in N}\rho_kx_{ik}+\delta_iF_i(\frac{y_i}{Y_i})-\kappa_i=0,\\
\frac{\partial\mathcal{L}}{\partial x_{E,ji}} &: -\rho_i+\beta_j\frac{\partial g_j}{\partial x_{E,ji}}=0,\\
\frac{\partial\mathcal{L}}{\partial y_i} &: \delta_iM_iF_i'(\frac{y_i}{Y_i})\frac{1}{Y_i}-\alpha_i=0,\\
\frac{\partial\mathcal{L}}{\partial M_{E,j}} &: -\beta_j+\sum_{i\in N}\kappa_i\zeta(j,i)=0.
\end{aligned}$$

Rearrange these

$$\begin{aligned}
\frac{\partial U}{\partial C_i} &= \rho_i,\\
\rho_i &= \delta_iM_iF_i'(\frac{y_i}{Y_i})\frac{y_i}{Y_i}\frac{1}{Y_i},\\
\rho_jM_i &=\alpha_iA_i\frac{\partial f_i}{\partial x_{ij}},\\
\kappa_i &= -\sum_{k\in N}\rho_kx_{ik}+\delta_iF_i(\frac{y_i}{Y_i}),\\
\rho_i &= \beta_j\frac{\partial g_j}{\partial x_{E,ji}},\\
\alpha_i &= \delta_iM_iF_i'(\frac{y_i}{Y_i})\frac{1}{Y_i},\\
\beta_j &= \sum_{i\in N}\kappa_i\zeta(j,i).
\end{aligned}$$

Now consider the equations that determine the decentralized equilibrium outcome, imposing that $\mu_i = \gamma_i = 1/\mu_i^Y$:

$$\frac{\partial U}{\partial C_i} = \frac{P_i}{P^Y}$$

$$mc_i A_i \frac{\partial f_i}{\partial x_{ij}} = P_j$$

$$\frac{p_i}{P_i} = F_i'(\frac{y_i}{Y_i})$$

$$P_{E,j} \frac{\partial g_j}{\partial x_{E,ji}} = P_i$$

$$\sum_{i \in N} \frac{\zeta(j,i) M_{E,j}}{M_i} \left[M_i p_i y_i - M_i \sum_{j \in \mathcal{N}} P_j x_{ij} \right] = \sum_{k \in \mathcal{N}} P_k x_{E,jk}$$

$$Y_i = C_i + \sum_{j \in \mathcal{N}} M_j x_{ji} + \sum_{j \in E} M_{E,j} x_{E,ji}$$

$$M_i = \sum_{j \in E} \zeta(j,i) M_{E,j},$$

where mc_i is the marginal cost of producer i. Note that the first-order conditions for the planning problem coincide with those that characterize the decentralized equilibrium. Specifically,

$$\frac{P_i}{P^Y} = \rho_i$$

$$\frac{mc_i}{P^Y} M_i = \alpha_i,$$

$$\frac{P_{E,j}}{P^Y} = \beta_j,$$

$$\delta_i = \gamma_i \frac{P_i Y_i}{P^Y},$$

$$\gamma_i^{-1} = M_i F_i'(\frac{y_i}{Y_i}) \frac{y_i}{Y_i},$$

$$\frac{1}{P^Y} \left[\gamma_i P_i Y_i \frac{1}{M_i} - \sum_{k \in N} P_i x_{ik} \right] = \kappa_i.$$

$$P_{E,j} = \sum_{i \in N} \left[\gamma_i P_i Y_i \frac{1}{M_i} - \sum_{k \in N} P_i x_{ik} \right] \zeta(j,i).$$

To verify these relationships, it helps to recognize that

$$p_i y_i = \gamma_i \frac{P_i Y_i}{M_i}.$$

■

Proof of Theorem 2. This follows from an application of the envelope theorem. ■

Lemma 4. *In equilibrium, the change in the mass of producers in each market is given by*

$$\mathrm{d}\log M = \widehat{\mathrm{d}\log\lambda}_{\pi} - \tilde{\zeta}'(\tilde{\zeta}\lambda_{\pi}\tilde{\zeta}')^{-1}\lambda_E \,\mathrm{d}\log P_E, \tag{23}$$

where $\mathrm{d}\log M$ *is the* $|\mathcal{N}| \times 1$ *vector of changes in masses of producers,* λ_E *is the* $|E| \times |E|$ *diagonal matrix of quasi-rents (expenditures on entry), and* $d\log P_E$ *is the* $|E| \times 1$ *vector of changes in entry prices.*[1] *Furthermore,*

Proof of Lemma 4. We assume that the rows of ζ are linearly independent, otherwise there are trivial entry types. Initialize the equilibrium where all M_E have been normalized to unity, we have the zero-profit conditions

$$\begin{aligned} \lambda_{E,i} &= \sum_{j\in N} \left(\frac{\zeta_{ij}}{\sum_{k\in E}\zeta_{kj}}\right)\lambda_{\pi_j}, \\ &= \sum_{j\in N} \tilde{\zeta}_{ij}\lambda_{\pi_j}, \end{aligned}$$

where $\tilde{\zeta}_{ij} = \zeta_{ij}M_{E,i}/\left(\sum_{k\in E}\zeta_{kj}M_{E,k}\right)$. Using the fact that

$$M_i = \sum_j \zeta_{ji} M_{E,j}. \tag{24}$$

loglinearize to get the zero-profit condition

$$\sum_j \tilde{\zeta}_{ij}\lambda_{\pi_j} d\log\lambda_{\pi_j} - \left(\sum_j \tilde{\zeta}_{ij}\lambda_{\pi_j}\right)\left(\sum_j \Omega^E_{ij} d\log P_j\right) = \sum_j \tilde{\zeta}_{ij}\lambda_{\pi_j} d\log M_j, \tag{25}$$

or in matrix notation, where λ_E and λ_π are diagonal matrices:

$$\tilde{\zeta}\lambda_\pi d\log\lambda_\pi - \lambda_E\Omega^E d\log P = \tilde{\zeta}\lambda_\pi\tilde{\zeta}' d\log M_E. \tag{26}$$

[1]The entry price $P_{E,j}$ of the jth entrant is the marginal cost associated with the production function in equation (2).

If ζ has linearly independent columns then

$$(\tilde{\zeta}\lambda_\pi\tilde{\zeta}')^{-1}\left(\tilde{\zeta}\lambda_\pi d\log\lambda_\pi - \lambda_E\Omega^E d\log P\right) = d\log M_E, \tag{27}$$

and

$$\tilde{\zeta}'(\tilde{\zeta}\lambda_\pi\tilde{\zeta}')^{-1}\left(\tilde{\zeta}\lambda_\pi d\log\lambda_\pi - \lambda_E\Omega^E d\log P\right) = d\log M. \tag{28}$$

From constant returns to scale, we know that $\Omega^E d\log P = d\log P_E$. ■

Proof of Theorem 3. The aggregation equation is

$$d\log Y = -\Omega_{(0,:)} d\log P. \tag{29}$$

Define $\varepsilon_i = 1$ if $i \in \mathcal{N}^{IRS}$ and $\gamma_i = 1$ if $i \in \mathcal{N}^{DRS}$. For each individual variety, we can write

$$\begin{aligned} d\log p_i &= d\log\mu_i + \sum_j (1-\pi_i)^{-1}\Omega^V_{ij} d\log P_j - d\log A_i + \frac{1-\varepsilon_i}{\varepsilon_i} d\log y_i, \\ &= d\log\mu_i + \sum_j \frac{\mu_i}{\varepsilon_i}\Omega^V_{ij} d\log P_j - d\log A_i + \frac{1-\varepsilon_i}{\varepsilon_i} d\log y_i. \end{aligned}$$

For the aggregated price, we have

$$\begin{aligned} d\log P_i &= d\log\mu^Y_i + d\log p_i - (\gamma_i - 1) d\log M_i \\ &= d\log\mu^Y_i + d\log\mu_i + \sum_j \frac{\mu_i}{\varepsilon_i}\Omega^V_{ij} d\log P_j - d\log A_i + \frac{1-\varepsilon_i}{\varepsilon_i} d\log y_i - (\gamma_i - 1) d\log M_i \\ &= d\log\mu^Y_i + d\log\mu_i + \sum_j \frac{\mu_i}{\varepsilon_i}\Omega^V_{ij} d\log P_j - d\log A_i \\ &+ \frac{1-\varepsilon_i}{\varepsilon_i}\left(d\log\lambda_i - d\log P_i - \gamma_i d\log M_i\right) - (\gamma_i - 1) d\log M_i \\ &= d\log\mu^Y_i + d\log\mu_i + \sum_j \frac{\mu_i}{\varepsilon_i}\Omega^V_{ij} d\log P_j - d\log A_i \\ &+ \frac{1-\varepsilon_i}{\varepsilon_i}\left(d\log\lambda_i - d\log P_i - d\log M_i\right) - \frac{1}{\varepsilon_i}(\gamma_i - 1) d\log M_i \end{aligned}$$

Using the fact that

$$d\log M = \widehat{d\log\lambda}_\pi - \tilde{\zeta}'\left(\tilde{\zeta}\lambda_\pi\tilde{\zeta}'\right)^{-1}\lambda_E\Omega^E d\log P, \tag{30}$$

we have

$$\begin{aligned} d\log P &= \varepsilon d\log \mu + \varepsilon d\log \mu^{Y} - \varepsilon d\log A \\ &+ (1-\varepsilon_i)\left(d\log\lambda - \widehat{d\log\lambda}_{\pi}\right) - (\gamma-1)\left(\widehat{d\log\lambda}_{\pi}\right) \\ &+ \mu\Omega^{V} d\log P + (\gamma-\varepsilon)\left(\tilde{\zeta}'\left(\tilde{\zeta}\lambda_{\pi}\tilde{\zeta}'\right)^{-1}\lambda_E\Omega^E d\log P\right) \end{aligned}$$

Letting

$$\Psi^{F} = \left(I - \varepsilon\mu\Omega^{V} - (\gamma-\varepsilon)\tilde{\zeta}'\left(\tilde{\zeta}\lambda_{\pi}\tilde{\zeta}'\right)^{-1}\lambda_E\Omega^E\right)^{-1} \tag{31}$$

gives

$$d\log P = \Psi^{F}\left(\varepsilon d\log\mu + \varepsilon d\log\mu^{Y} - \varepsilon d\log A + (1-\varepsilon)\left(d\log\lambda - \widehat{d\log\lambda}_{\pi}\right) - (\gamma-1)\left(\widehat{d\log\lambda}_{\pi}\right)\right), \tag{32}$$

which can be rearranged, using $d\log Y = -\Omega_{(0,:)} d\log P$, to give desired result. ■

Proof of Proposition 1. The proof for this is the same as that of Theorem 3. ■

Proof of Proposition 2. Note that

$$\lambda_{\pi_i} = \left(1 - \frac{\varepsilon_i}{\mu_i}\right)\frac{\lambda_i^B}{\mu_i^Y}.$$

$$\begin{aligned} d\log\lambda_{\pi_i} &= d\log\lambda_i^B - d\log\mu_i^Y + d\log\pi_i \\ &= d\log\lambda_i^B - d\log\mu_i^Y + \frac{1-\pi_i}{\pi_i} d\log\mu_i. \end{aligned}$$

The free entry condition ensures that

$$\lambda_E = diag\left(M_E\right)\zeta diag(M)^{-1}\pi\lambda^B, \tag{33}$$

So,

$$\begin{aligned} \lambda^{B'} &= \lambda^{B'}\Omega^V + \left(\lambda_E\right)'\Omega^E, \\ &= \lambda^{B'}\Omega^V + \lambda^{B'}\pi diag(M)^{-1}\zeta' diag\left(M_E\right)\Omega^E. \end{aligned}$$

Therefore,

$$d\lambda^{B'} = \lambda^{B'} d\Omega^V + \lambda^{B'} d\pi diag(M)^{-1}\zeta' diag\left(M_E\right)\Omega^E.$$

$$
\begin{aligned}
&- \lambda^{B'} \pi diag(M)^{-1} diag(d\log M) \zeta' diag\left(M_E\right) \Omega^E \\
&+ \lambda^{B'} \pi diag(M)^{-1} \zeta' diag\left(M_E\right) d\Omega^E \\
&+ \lambda^{B'} \pi diag(M)^{-1} \zeta' diag\left(M_E\right) diag(d\log M_E) d\Omega^E \\
&+ d\lambda^{B'} \left(\Omega^V + \pi diag(M)^{-1} \zeta' \Omega^E\right), \\
&= \lambda^{B'} \left(d\Omega^V + d\pi diag(M)^{-1} \zeta' diag\left(M_E\right) \Omega^E\right) \Psi^B \\
&- \lambda^{B'} \left(\pi diag(M)^{-1} diag(d\log M) \zeta' diag\left(M_E\right) \Omega^E\right) \Psi^B \\
&+ \lambda^{B'} \left(\pi diag(M)^{-1} \zeta' diag\left(M_E\right) diag(d\log M_E) d\Omega^E\right) \Psi^B \\
&+ \lambda^{B'} \left(\pi diag(M)^{-1} \zeta' diag\left(M_E\right) d\Omega^E\right),
\end{aligned} \tag{34}
$$

where, using the fact that in the initial equilibrium $\zeta diag(M)^{-1} = \tilde{\zeta}$

$$
\begin{aligned}
\Psi^B &= \left(I - \Omega^V - \pi diag(M)^{-1} \zeta' diag\left(M_E\right) \Omega^E\right), \\
&= \left(I - \Omega^V - \pi \tilde{\zeta}' \Omega^E\right).
\end{aligned}
$$

Using the fact that

$$
d\Omega^V_{ij} = -\Omega^V_{ij} d\log\left(\mu_i \mu_i^Y\right) + \mu_i^{-1}(1-\theta_i) Cov_i\left(d\log P, I_{(j)}\right), \tag{35}
$$

we can rewrite (34) as

$$
\begin{aligned}
d\lambda_i^B =& -\sum_{m\in N} \lambda_m^B \sum_{k\in N} \Omega^V_{mk} \Psi^B_{ki} d\log\left(\mu_m \mu_m^Y\right) + \sum_m \lambda_m^B \mu_m^{-1}(1-\theta_m) Cov_m\left(d\log P, \Psi^B_{(i)}\right) \\
&+ \sum_{j\in E} \sum_{m\in N} \sum_{k\in N} \lambda_m^B \varepsilon_m \frac{d\log \mu_m}{\mu_k \mu_k^Y} \tilde{\zeta}_{jm} \Omega^E_{jk} \Psi^B_{ki} \\
&- \sum_{k\in N} \sum_{m\in N} \sum_{j\in E} \lambda_m^B \pi_m \tilde{\zeta}_{jm} \Omega^E_{jk} \left(d\log M_m + d\log \mu^Y\right) \Psi^B_{ki} \\
&+ \sum_{k\in N} \sum_{m\in N} \lambda_m^B \sum_{j\in E} \pi_m \tilde{\zeta}_{jm} \Omega^E_{jk} d\log M_{E,j} \Psi^B_{ki},
\end{aligned}
$$

where we use the fact that we have assumed (without loss of generality) that Ω^E is degenerate.

In Appendix C, we introduce the notion of non-overlapping entry and show that we can impose this without loss of generality. Under non-overlapping entry, we use the following identity

Lemma 5. *Under non-overlapping entry, the following identity holds:*

$$\sum_{k\in N}\sum_{m\in N}\sum_{j\in E}\lambda_m^B\pi_m\tilde{\zeta}_{jm}\Omega_{jk}^E d\log M_m\Psi_{ki}^B \quad = \quad \sum_{k\in N}\sum_{m\in N}\sum_{j\in E}\lambda_m^B\pi_m\tilde{\zeta}_{jm}\Omega_{jk}^E d\log M_{E,j}\Psi_{ki}^B. \quad (36)$$

Proof. Rearrange the left-hand side to be:

$$\begin{aligned}\sum_{k\in N}\sum_{m\in N}\sum_{j\in E}\tilde{\zeta}_{jm}\lambda_{\pi_m}\Omega_{jk}^E\left(d\log M_m - d\log M_{E,j}\right)\Psi_{ki}^B &= \sum_{k\in N}\sum_{j\in E}\Omega_{jk}^E\Psi_{ki}^B\sum_{m\in N}\tilde{\zeta}_{jm}\lambda_{\pi_m}\left(d\log M_m - d\log M_{E,j}\right)\\ &= \sum_{k\in N}\sum_{j\in E}\Omega_{jk}^E\Psi_{ki}^B\left(\sum_{m\in N}\tilde{\zeta}_{jm}\lambda_{\pi_m}\left(d\log M_m\right) - \lambda_{E,j}d\log M_{E,j}\right)\\ &= \sum_{k\in N}\sum_{j\in E}\Omega_{jk}^E\Psi_{ki}^B\left(\sum_{m\in N}\tilde{\zeta}_{jm}\lambda_{\pi_m}\left(d\log M_m\right) - \lambda_{E,j}d\log M_{E,j}\right).\end{aligned}$$

The free-entry condition is

$$P_{E,j} = \sum_{k\in N}\zeta_{jk}\lambda_{\pi_k}\frac{1}{M_k}. \quad (37)$$

$$\begin{aligned}\lambda_{E,j}d\log P_{E,j} &= \sum_{k\in N}\tilde{\zeta}_{jk}\lambda_{\pi_k}d\log\lambda_{\pi_k} - \sum_{k\in N}\tilde{\zeta}_{jk}\lambda_{\pi_k}d\log M_k,\\ \sum_{k\in N}\tilde{\zeta}_{jk}\lambda_{\pi_k}d\log M_k &= \sum_{k\in N}\tilde{\zeta}_{jk}\lambda_{\pi_k}d\log\lambda_{\pi_k} - \lambda_{E,j}d\log P_{E,j}\\ \tilde{\zeta}\lambda_\pi d\log M &= \tilde{\zeta}\lambda_\pi d\log\lambda_\pi - \lambda_E d\log P_E.\end{aligned}$$

On the other hand,

$$\lambda_E d\log\lambda_E = \lambda_E d\log M_E + \lambda_E d\log P_E. \quad (38)$$

Finally, note that, free entry requires that

$$\begin{aligned}\lambda_E &= \tilde{\zeta}\lambda_\pi,\\ \lambda_E d\log\lambda_E &= \tilde{\zeta}\lambda_\pi d\log\lambda_\pi + \tilde{\zeta}d\log\tilde{\zeta}\lambda_\pi.\end{aligned}$$

If there is non-overlapping entry, then

$$d\log\tilde{\zeta} = 0. \quad (39)$$

Hence,

$$\lambda_E d\log M_E = \lambda_E d\log\lambda_E - \lambda_E d\log P_E$$

$$= \tilde{\zeta}\lambda_\pi d\log\lambda_\pi - \lambda_E d\log P_E$$
$$= \tilde{\zeta}\lambda_\pi d\log M.$$

Therefore,

$$\sum_{k\in N}\sum_{j\in E}\Omega^E_{jk}\Psi^B_{ki}\left(\sum_{m\in N}\tilde{\zeta}_{jm}\lambda_{\pi_m}\left(d\log M_m\right) - \lambda_{E,j}d\log M_{E,j}\right) = 0, \tag{40}$$

as needed. In general,

$$\begin{aligned}
\lambda_E &= \tilde{\zeta}\lambda_\pi, \\
\lambda_E d\log\lambda_E &= \tilde{\zeta}\lambda_\pi d\log\lambda_\pi + d\tilde{\zeta}\lambda_\pi \\
&= \tilde{\zeta}\lambda_\pi d\log\lambda_\pi + d\log M_E\tilde{\zeta}\lambda_\pi - \tilde{\zeta}d\log M\lambda_\pi
\end{aligned}$$

Hence

$$\begin{aligned}
\lambda_E d\log M_E &= \lambda_E d\log\lambda_E - \lambda_E d\log P_E \\
&= \tilde{\zeta}\lambda_\pi d\log\lambda_\pi + d\log M_E\tilde{\zeta}\lambda_\pi - \tilde{\zeta}d\log M\lambda_\pi - \lambda_E d\log P_E \\
&= \tilde{\zeta}\lambda_\pi d\log M + d\log M_E\tilde{\zeta}\lambda_\pi - \tilde{\zeta}d\log M\lambda_\pi \\
&= d\log M_E\tilde{\zeta}\lambda_\pi
\end{aligned}$$

In other words,

$$\sum_{k\in N}\sum_{j\in E}\Omega^E_{jk}\Psi^B_{ki}\left(\sum_{m\in N}\tilde{\zeta}_{jm}\lambda_{\pi_m}\left(d\log M_m - d\log M_{E,j}\right)\right) = 0. \tag{41}$$

Simplify it a bit

$$\begin{aligned}
\tilde{\zeta}\lambda_\pi d\log M &= \tilde{\zeta}\lambda_\pi\tilde{\zeta}'(\tilde{\zeta}\lambda_\pi\tilde{\zeta}')^{-1}\left(\tilde{\zeta}\lambda_\pi d\log\lambda_\pi - \lambda_E\Omega^E d\log P\right) \\
&= \tilde{\zeta}\lambda_\pi d\log\lambda_\pi - \lambda_E\Omega^E d\log P
\end{aligned}$$

$$\begin{aligned}
\lambda_E d\log M_E &= \lambda_E\left((\tilde{\zeta}\lambda_\pi\tilde{\zeta}')^{-1}\left(\tilde{\zeta}\lambda_\pi d\log\lambda_\pi - \lambda_E\Omega^F d\log P\right)\right), \\
&= diag(\tilde{\zeta}\lambda_\pi\mathbf{1})(\tilde{\zeta}\lambda_\pi\tilde{\zeta}')^{-1}\left(\tilde{\zeta}\lambda_\pi d\log\lambda_\pi - \lambda_E\Omega^F d\log P\right).
\end{aligned}$$

Hence

$$\begin{aligned}\tilde{\zeta}\lambda_\pi d\log M - \lambda_E d\log M_E &= \tilde{\zeta}\lambda_\pi d\log\lambda_\pi \\ &- diag(\tilde{\zeta}\lambda_\pi \mathbf{1})\Omega^F d\log P \\ &- diag(\tilde{\zeta}\lambda_\pi \mathbf{1})(\tilde{\zeta}\lambda_\pi\tilde{\zeta}')^{-1}\left(\tilde{\zeta}\lambda_\pi d\log\lambda_\pi - diag(\tilde{\zeta}\lambda_\pi\mathbf{1})\Omega^F d\log P\right) \\ &= \left(I_{E\times E} - diag(\tilde{\zeta}\lambda_\pi\mathbf{1})(\tilde{\zeta}\lambda_\pi\tilde{\zeta}')^{-1}\right)\left(\tilde{\zeta}\lambda_\pi d\log\lambda_\pi - diag(\tilde{\zeta}\lambda_\pi\mathbf{1})\Omega^F d\log P\right)\end{aligned}$$

where we use the fact that

$$(\tilde{\zeta}\lambda_\pi\tilde{\zeta}')^{-1}\left(\tilde{\zeta}\lambda_\pi d\log\lambda_\pi - \lambda_E\Omega^E d\log P\right) = d\log M_E, \tag{42}$$

and

$$\tilde{\zeta}'(\tilde{\zeta}\lambda_\pi\tilde{\zeta}')^{-1}\left(\tilde{\zeta}\lambda_\pi d\log\lambda_\pi - \lambda_E\Omega^E d\log P\right) = d\log M. \tag{43}$$

Hence, in general we have

$$\sum_{k\in N}\sum_{j\in E}\Omega^E_{jk}\Psi^B_{ki}\left(\sum_{m\in N}\tilde{\zeta}_{jm}\lambda_{\pi_m}\left(d\log M_m\right) - \lambda_{E,j}d\log M_{E,j}\right) = \Big[\left(I_{E\times E} - diag(\tilde{\zeta}\lambda_\pi\mathbf{1})(\tilde{\zeta}\lambda_\pi\tilde{\zeta}')^{-1}\right)$$
$$\left(\tilde{\zeta}\lambda_\pi d\log\lambda_\pi - diag(\tilde{\zeta}\lambda_\pi\mathbf{1})\Omega^F d\log P\right)\Big]'\,\Omega^E\Psi^B. \tag{44}$$

■

Having defined

$$d\log\hat{P} = \tilde{\zeta}'(\tilde{\zeta}\lambda_\pi\tilde{\zeta}')^{-1}\lambda_E\Omega^F d\log P, \tag{45}$$

with the aid of the lemma above, if we have non-overlapping entry, we get the simpler expressions

$$\begin{aligned}d\lambda^B_i = &-\sum_{m\in N}\lambda^B_m\sum_{k\in N}\Omega^V_{mk}\Psi^B_{ki}d\log\left(\mu_m\mu^Y_m\right) + \sum_{j\in E}\sum_{m\in N}\sum_{k\in N}\lambda^B_m\varepsilon_m\frac{d\log\left(\mu_m\right)}{\mu_m\mu^Y_m}\tilde{\zeta}_{jm}\Omega^E_{jk}\Psi^B_{ki} \\ &-\sum_{k\in N}\sum_{m\in N}\sum_{j\in E}\lambda^B_m\pi_m\tilde{\zeta}_{jm}\Omega^E_{jk}d\log\mu^Y_m\Psi^B_{ki} + \sum_m\frac{\lambda^B_m}{\mu^Y_m}(1-\theta_m)Cov_{\tilde{\Omega}^{V,m}}\left(d\log P, \Psi^B_{(i)}\right).\end{aligned}$$

■

Proof of Proposition 3. We start with

$$d\log Y = \sum_i b_i d\log C_i$$

$$C_i d\log C_i = Y_i d\log Y_i - \sum_{j\in N} x_{ji} d\log x_{ji} M_j - \sum_{j\in N} x_{ji} M_j d\log M_j - \sum_{j\in E} x_{E,ji} d\log x_{E,ji}$$

$$P_i C_i d\log C_i = P_i Y_i d\log Y_i - \sum_{j\in N} P_i x_{ji} M_j d\log x_{ji} - \sum_{j\in N} P_i x_{ji} M_j d\log M_j - \sum_{j\in E} P_i x_{E,ji} d\log x_{E,ji}$$

$$b_i d\log C_i = \lambda_i^B d\log Y_i - \sum_{j\in N} \lambda_j^B \frac{P_i x_{ji}}{P_j Y_j} M_j \left(d\log x_{ji} + d\log M_j\right) - \sum_{j\in E} \lambda_j^E \frac{P_i x_{E,ji}}{\lambda_j^E} d\log x_{E,ji},$$

$$d\log Y = \sum_{i\in N}\left(\lambda_i^B d\log Y_i - \sum_{j\in N} \lambda_j^B \frac{P_i x_{ji}}{P_j Y_j} M_j \left(d\log x_{ji} + d\log M_j\right) - \sum_{j\in E} \lambda_j^E \frac{P_i x_{E,ji}}{\lambda_j^E} d\log x_{E,ji}\right),$$

$$= \sum_{i\in N}\left(\lambda_i^B d\log Y_i - \sum_{j\in N} \lambda_i^B \frac{P_j x_{ij}}{P_i Y_i} M_i \left(d\log x_{ij} + d\log M_i\right) - \sum_{j\in E} \lambda_j^E \frac{P_i x_{E,ji}}{\lambda_j^E} d\log x_{E,ji}\right),$$

We also have

$$P_i Y_i = \mu_i^Y M_i p_i^y y_i.$$

$$d\log Y_i = d\log y_i + \gamma_i d\log M_i. \tag{46}$$

Let

$$q_i = f_i(x_{ij}), \quad y_i = q_i^{\varepsilon_i}. \tag{47}$$

Let p_i^qbe the marginal cost of producing $q_{i.}$ Hence

$$\frac{\partial \log q_i}{\partial \log x_{ij}} = \frac{P_j x_{ij}}{p_i^q q_i} = \mu_i^Y \mu_i \frac{1}{\varepsilon_i} M_i \frac{P_j x_{ij}}{P_i Y_i} = \mu_i^Y \mu_i \frac{1}{\varepsilon_i} \Omega_{ij}^V, \tag{48}$$

Furthremore,

$$d\log q_i = \sum_j \frac{p_j x_{ij}}{p_i q_i} d\log x_{ij}. \tag{49}$$

So,

$$\begin{aligned} d\log Y_i &= d\log y_i + \gamma_i d\log M_i \\ &= \gamma_i d\log M_i + \varepsilon_i \sum_j \frac{\partial \log q_i}{\partial \log x_{ij}} d\log x_{ij} \\ &= \gamma_i d\log M_i + \mu_i^Y \mu_i^y \sum_j \Omega_{ij}^V d\log x_{ij} \end{aligned}$$

$$\sum_j \Omega_{ij}^V d\log x_{ij} + \sum_j \Omega_{ij}^V d\log M_i = \frac{1}{\mu_i \mu_i^Y} d\log Y_i - \frac{\gamma_i}{\mu_i \mu_i^Y} d\log M_i + \frac{\varepsilon_i}{\mu_i \mu_i^Y} d\log M_i$$

We can write

$$
\begin{aligned}
d\log Y &= \sum_{i\in N}\left(\lambda_i^B d\log Y_i - \sum_{j\in N}\lambda_i^B \frac{P_j x_{ij}}{P_i Y_i} M_i\left(d\log x_{ij} + d\log M_i\right) - \sum_{j\in E}\lambda_j^E \frac{P_i x_{E,ji}}{\lambda_j^E} d\log x_{E,ji}\right),\\
&= \sum_{i\in N}\left(\lambda_i^B d\log Y_i - \lambda_i^B \sum_{j\in N}\Omega_{ij}^V\left(d\log x_{ij} + d\log M_i\right) - \sum_{j\in E}\lambda_j^E \frac{P_i x_{E,ji}}{\lambda_j^E} d\log x_{E,ji}\right),\\
&= \sum_{i\in N}\left(\lambda_i^B d\log Y_i - \lambda_i^B \frac{1}{\mu_i^Y \mu_i}\left[d\log Y_i - \gamma_i d\log M_i + \varepsilon_i d\log M_i\right] - \sum_{j\in E}\Omega_{ji}^E \lambda_{E,j} d\log x_{E,ji}\right),\\
&= \sum_{i\in N}\left(\lambda_i^B\left(1 - \frac{1}{\mu_i^Y\mu_i}\right) d\log Y_i\right) + \sum_{i\in N}\lambda_i^B \frac{\gamma_i - \varepsilon_i}{\mu_i^Y \mu_i} d\log M_i - \sum_{j\in E}\sum_{i\in N}\Omega_{ji}^E \lambda_{E,j} d\log x_{E,ji},\\
&= \sum_{i\in N}\left(\lambda_i^B\left(1 - \frac{1}{\mu_i^Y\mu_i}\right) d\log Y_i\right) + \sum_{i\in N}\lambda_i^B \frac{\gamma_i - \varepsilon_i}{\mu_i^Y \mu_i} d\log M_i - \sum_{j\in E}\lambda_{E,j} d\log M_{E,j},\\
&= \sum_{i\in N}\left(\lambda_i^B\left(1 - \frac{1}{\mu_i^Y\mu_i}\right) d\log Y_i\right) + \sum_{i\in N}\lambda_i^B \frac{\gamma_i - \varepsilon_i}{\mu_i^Y \mu_i} \sum_{j\in E}\tilde{\zeta}_{ij} d\log M_{E,j} - \sum_{j\in E}\lambda_{E,j} d\log M_{E,j}.
\end{aligned}
$$

Finally, note that

$$
\lambda_{E,j} = \sum_i \frac{\lambda_i^B}{\mu_i^Y}\left(1 - \frac{\varepsilon_i}{\mu_i}\right)\tilde{\zeta}_{ij}
$$

Hence,

$$
\begin{aligned}
d\log Y &= \sum_{i\in N}\left(\lambda_i^B\left(1 - \frac{1}{\mu_i^Y\mu_i}\right) d\log Y_i\right) + \sum_{i\in N}\lambda_i^B \frac{\gamma_i - \varepsilon_i}{\mu_i^Y \mu_i} \sum_{j\in E}\tilde{\zeta}_{ij} d\log M_{E,j} - \sum_{j\in E}\sum_i \frac{\lambda_i^B}{\mu_i^Y}\left(1 - \frac{\varepsilon_i}{\mu_i}\right)\tilde{\zeta}_{ij} d\log M_{E,j},\\
&= \sum_{i\in N}\left(\lambda_i^B\left(1 - \frac{1}{\mu_i^Y\mu_i}\right) d\log Y_i\right) + \sum_{i\in N}\frac{\lambda_i^B}{\mu_i^Y}\left(\frac{\gamma_i - \varepsilon_i}{\mu_i} - 1 + \frac{\varepsilon_i}{\mu_i}\right)\sum_{j\in E}\tilde{\zeta}_{ij} d\log M_{E,j},\\
&= \sum_{i\in N}\left(\lambda_i^B\left(1 - \frac{1}{\mu_i^Y\mu_i}\right) d\log Y_i\right) + \sum_{i\in N}\frac{\lambda_i^B}{\mu_i^Y}\left(\frac{\gamma_i}{\mu_i} - 1\right)\sum_{j\in E}\tilde{\zeta}_{ij} d\log M_{E,j}.
\end{aligned}
$$

Diffrentiate this expression a second time with respect to $\log\mu$and $\log\mu^Y$and evaluate it at the efficient point to get

$$
d^2\log Y = \frac{1}{2}\left[\sum_{i\in N}\lambda_i^B d\log Y_i d\log\left(\mu_i\mu_i^Y\right) - \sum_{i\in N}\lambda_i^B\gamma_i\sum_{j\in E}\tilde{\zeta}_{ij} d\log\mu_i d\log M_{E,j}\right].
$$

Now use the fact that and use the fact that and

$$d\log M_i = \sum_{j\in E} \tilde{\zeta}_{ij} d\log M_{E,j}. \tag{50}$$

to get

$$\begin{aligned}
d^2 \log Y &= \frac{1}{2}\left[\sum_{i\in N} \lambda_i^B d\log Y_i d\log\left(\mu_i \mu_i^Y\right) - \sum_{i\in N} \lambda_i^B \gamma_i d\log\mu_i d\log M_i\right] \\
&= \frac{1}{2}\left[\sum_{i\in N} \lambda_i^B \left[d\log y_i\right] d\log\left(\mu_i \mu_i^Y\right) + \sum_{i\in N} \lambda_i^B \gamma_i d\log M_i d\log\left(\mu_i \mu_i^Y\right) - \sum_{i\in N} \lambda_i^B \gamma_i d\log\mu_i d\log M_i\right] \\
&= \frac{1}{2}\left[\sum_{i\in N} \lambda_i^B \left[d\log y_i\right] d\log\left(\mu_i \mu_i^Y\right) + \sum_{i\in N} \lambda_i^B \gamma_i d\log M_i d\log\mu_i^Y\right].
\end{aligned}$$

■

Proof of Proposition 4. We assume there is one primary factor with no incumbents, no input-output in entry costs. For a model with entry in sectors, we can assume away within-industry heterogeneity momentarily. Therefore, we can assume entry is fully directed. We use the deadweight loss triangles formula, along with the fact that for each $i \in \mathcal{N}$

$$\mathrm{d}\log Y_i = \mathrm{d}\log\lambda_i^B - \mathrm{d}\log P_i.$$

So,

$$d\log\lambda_l^B = \sum_k \left(\delta_{lk} - \frac{\lambda_k^B}{\lambda_l^B}\Psi_{kl}^B\right) d\log\mu_k^q - \sum_j \frac{\lambda_j}{\lambda_l}(\theta_j - 1) Cov_j(d\log P, \Psi_{(l)}^B), \tag{51}$$

where δ_{lk} is Kronecker's delta, and

$$d\log\lambda_{\pi_i} = d\log\lambda_i^B + \left(\frac{1}{1-\varepsilon_i} - 1\right) d\log\mu_i, \tag{52}$$

$$\begin{aligned}
d\log P &= \Psi^F\left(\frac{\varepsilon}{\gamma} d\log\mu\right) + \Psi^F\left(1 - \varepsilon\left(d\log\lambda - d\log\hat{\lambda}_\pi\right)\right), \\
&= \Psi^F(\varepsilon) d\log\mu - \Psi^F(\varepsilon) d\log\mu = 0.
\end{aligned}$$

Hence

$$d\log\lambda_l^B = \sum_k \left(\delta_{lk} - \frac{\lambda_k^B}{\lambda_l^B}\Psi_{kl}^B\right) d\log\mu_k^q, \tag{53}$$

Furthermore, letting Λ denote labor's share of income

$$\begin{aligned} d\log M_E &= d\log \lambda_\pi - d\log \Lambda, \\ &= d\log \lambda_i^B + \left(\frac{1}{1-\varepsilon_i} - 1\right) d\log \mu_i \end{aligned}$$

$$\sum_l \lambda_l^B \left(d\log \lambda_l^B - d\log p_l\right) d\log \mu_l = \sum_l \sum_k \left(\lambda_l^B \delta_{lk} - \lambda_k^B \Psi_{kl}^B\right) d\log \mu_k d\log \mu_l$$

Next

$$\sum_{j\in E} \sum_{i\in N} \frac{\lambda_i^B \tilde{\zeta}_{ij}}{\mu_i^Y \gamma_i} d\log\left(\mu_i^y \mu_i^q\right) d\log M_{E,j} = \sum_i \lambda_i^B d\log \mu_i d\log \lambda_i^B + \sum_i \lambda_i^B \left(\frac{1}{1-\varepsilon_i} - 1\right) d\log \mu_i d\log \mu_i$$

Combining everything gives

$$\mathcal{L} = \sum_l \sum_k \left(\lambda_l^B \delta_{lk} - \lambda_k^B \Psi_{kl}^B\right) d\log \mu_k d\log \mu_l - \sum_i \lambda_i d\log \mu_i d\log \lambda_i^B - \sum_i \lambda_i^B \left(\frac{1}{1-\varepsilon_i} - 1\right) d\log \mu_i d\log \mu_i.$$

$$d\log \lambda_i^B = -\frac{1}{\lambda_i^B} \sum_j \left(\lambda_j^B \Psi_{ji}^B - \lambda_j^B \delta_{ij}\right) d\log \mu_j, \tag{54}$$

Or

$$\begin{aligned} \mathcal{L} &= \sum_k \lambda_k^B \sum_l \left(\delta_{lk} - \Psi_{kl}^B\right) d\log \mu_k d\log \mu_l + \sum_j \lambda_j^B \sum_i \left(\Psi_{ji}^B - \delta_{ij}\right) d\log \mu_j d\log \mu_i \\ &- \sum_i \lambda_i^B \left(\frac{1}{1-\varepsilon_i} - 1\right) d\log \mu_i d\log \mu_i, \\ &= -\sum_i \lambda_i^B \left(\frac{1}{1-\varepsilon_i} - 1\right) d\log \mu_i d\log \mu_i. \end{aligned}$$

This is the loss function for a model with homogeneous sectors.

To extend this into a sectoral model with within-sector heterogeneity, consider the isomorphic sectoral model. We know that

$$d\log Y = \frac{d\log Y}{d\log A} d\log A + \frac{d\log Y}{d\log \mu} d\log \mu \tag{55}$$

$$\frac{1}{2} d^2 \log Y = \frac{1}{2} d\log A' \frac{d^2 \log Y}{d\log A^2} d\log A + \frac{d\log Y}{d\log A} d^2 \log A + \frac{1}{2} d\log \mu' \frac{d^2 \log Y}{d\log \mu^2} d\log \mu + \frac{d\log Y}{d\log \mu} d^2 \log \mu \tag{56}$$

At the efficient point, $d \log A = 0$ and $d \log Y / d \log \mu = 0$,

$$-\mathcal{L} = \frac{1}{2} \frac{d \log Y}{d \log A} d^2 \log A + \frac{1}{2} d \log \mu' \frac{d^2 \log Y}{d \log \mu^2} d \log \mu$$

where, from the proof of the previous proposition, we know that

$$d^2 \log A_k = -\frac{1}{2} \frac{1}{1-\varepsilon} Var_{\delta_k} \left(d \log \mu_{(k)} \right). \tag{57}$$

Finally, recall note that at the efficient point, from Hulten's theorem, $d \log Y / d \log A = \lambda^B(\varepsilon)$, so we get

$$d^2 \log Y = -\frac{1}{2} \sum_I \lambda_I^B \left(\frac{1}{1-\varepsilon_I} - 1 \right) Var_{\delta_I} \left(d \log \mu_{(I)} \right) - \frac{1}{2} \sum_I \lambda_I^B \left(\frac{1}{1-\varepsilon_I} - 1 \right) E_{\delta_I} \left(d \log \mu_{(I)} \right)^2$$

■

Appendix C The Role of Reallocation

To see how Theorem 3 can be decomposed into technical and allocative efficiency, without loss of generality, we impose the follow assumption.

Assumption 3. Entry is *non-overlapping*. That is, for each $i \in \mathcal{N}$, there is at most one entrant type $j \in E$ that can produce product i: $\zeta(j, i) \neq 0$.[2]

Theorem 3 provides an interpretable decomposition of changes in output into changes in technical and allocative efficiency along the lines of Baqaee and Farhi (2019a). To see this, let $\mathcal{X}$ denote the $(|\mathcal{N}| + |E|) \times |\mathcal{N}|$ *allocation matrix* of the economy, where $\mathcal{X}_{ij}$ records the fraction of good j used by a producer or entrant $i \in \mathcal{N} + E$. Together with the vector of productivity shifters A, the allocation matrix pins down the whole allocation, and hence aggregate output $Y(A, \mathcal{X})$.

In particular, equilibrium aggregate output is obtained by using the equilibrium allocation matrix $\mathcal{X}(A, \mu)$ where μ is the vector of markups/wedges. Changes in equilibrium

[2]To see why we can impose this without loss of generality, consider a situation where entrants 1 and 2 enter into the same market, so that $Y = F(My)$ and $M = M_{E,1} + M_{E,2}$. To turn this into a model with non-overlapping entry, create two fictitious markets $Y_i = F(M_i y_i)$ with non-overlapping entry $M_i = M_{E,i}$ for $i \in \{1, 2\}$. Now create a third fictitious market, with no entry, where Y_3 aggregates Y_1 and Y_1 in the same way as F. Since $Y = Y_3$, we have recast a model with overlapping entry into an equivalent model with non-overlapping entry. We impose this assumption throughout.

aggregate output in response to shocks can therefore be written, in matrix notation, as

$$\mathrm{d}\log Y = \underbrace{\frac{\partial \log Y}{\partial \log A}\mathrm{d}\log A}_{\Delta\text{Technical Efficiency}} + \underbrace{\frac{\partial \log Y}{\partial \mathcal{X}}\mathrm{d}\mathcal{X}}_{\Delta\text{Allocative Efficiecny}},$$

where the first term is the direct effect of changes in technology, holding the allocation of resources constant, and the second term is the indirect effect of equilibrium reallocations

$$\mathrm{d}\mathcal{X} = \frac{\partial \mathcal{X}}{\partial \log A}\mathrm{d}\log A + \frac{\partial \mathcal{X}}{\partial \log \mu}\mathrm{d}\log \mu.$$

Proposition 7 breaks Theorem 3 into two components.

Proposition 7 (Decomposition with Inefficiencies)**.** *In response to shocks* $(\mathrm{d}\log A, \mathrm{d}\log \mu)$*, changes into aggregate output can be decomposed in changes in technical efficiency*

$$\frac{\partial \log Y}{\partial \log A}\mathrm{d}\log A = \sum_{i\in\mathcal{N}} \lambda_i^F \,\mathrm{d}\log A_i,$$

and changes in allocative efficiency

$$\begin{aligned}\frac{\partial \log Y}{\partial \mathcal{X}}\mathrm{d}\mathcal{X} = &-\sum_{i\in\mathcal{N}^{IRS}} \lambda_i^F d\log \mu_i - \sum_{i\in\mathcal{N}^{DRS}} \lambda_i^F \varepsilon_i d\log \mu_i \\ &- \sum_{i\in\mathcal{N}^{DRS}} \lambda_i^F(1-\varepsilon_i)\left(d\log \lambda_i^B - \widehat{\mathrm{d}\log\lambda}_{\pi,i}\right) + \sum_{i\in\mathcal{N}^{IRS}} \lambda_i^F(\gamma_i - 1)\,\widehat{\mathrm{d}\log\lambda}_{\pi,i}.\end{aligned} \tag{58}$$

Changes in technical efficiency are a Hulten-like weighted sum of changes in productivities. The weights are forward Domar weights rather than traditional Domar weights. This is because when the allocation of resources is kept constant, productivity shocks are pushed forward through supply chains to the household, and the household's exposure in prices Ψ_{0i}^F to each good i is given by λ_i^F not λ_i^B.

Changes in allocative efficiency can be traced back to reductions in prices (shares) of specific fixed factors associated with individual producers and with entry. Focus on productivity shocks for simplicity, so that the first line of (58) is zero. This leaves two terms on the second line.

The first term depends on decreasing internal returns to scale $1-\varepsilon_i$. When $\mathrm{d}\log\lambda_i^{\mathrm{B}} - \widehat{\mathrm{d}\log\lambda}_{\pi,i} > 0$, this means that individual producers in market i are scaling up and running into diminishing returns. This raises the shadow price of their producer-specific fixed factor and contributes negatively to changes in allocative efficiency in proportion to the forward Domar weight $\lambda_i^F(1-\varepsilon_i)$ of these specific fixed factors.[3,4] When $\mathrm{d}\log\lambda_i^{\mathrm{B}} - \widehat{\mathrm{d}\log\lambda}_{\pi,i} = 0$, decreasing returns to scale do not matter since adjustments in market size are taking place along the extensive margin (individual producers are not change their scale).

The second term depends on consumer surplus $\gamma_i - 1$. When $\widehat{\mathrm{d}\log\lambda}_{\pi,i} > 0$, this means that entry is increasing in market i and triggering external economies from love of variety. This reduces the (negative) shadow price of the specific fixed factor associated with entry and contributes positively to changes in allocative efficiency in proportion to the forward Domar weight $\lambda_i^F(\gamma_i - 1)$ of these specific fixed factors.

Improvements in allocative efficiency can be measured by a forward-weighted sum of reductions in the shadow prices of fixed factors. Beneficial equilibrium reallocations, by using more resources more efficiently, reduce the shadow prices of fixed factors on balance across markets. This can only occur when the economy is inefficient. When the economy is efficient, reductions in the shadow prices of some specific fixed factors are exactly compensated by increases in others.

Corollary 1 (Decomposition under Efficiency)**.** *In the marginal-cost pricing equilibrium, as long as $\varepsilon_i, \gamma_i < 1$ for all $i \in \mathcal{N}$, changes in technical and allocative efficiency are given by*[5]

$$\frac{\partial \log Y}{\partial \log A}\,\mathrm{d}\log A = \sum_{i\in\mathcal{N}} \lambda_i^F\,\mathrm{d}\log A_i \qquad \textit{and} \qquad \frac{\partial \log Y}{\partial \mathcal{X}}\,\mathrm{d}\mathcal{X} = 0,$$

with $\lambda_i^F = \lambda_i^B$.

In the efficient benchmark, technology shocks only have direct effects and not indirect reallocation effects. Of course, this does not mean that reallocations do not occur in efficient models, but merely that their impact is irrelevant to a first order.

Appendix D Beyond CES

Following Baqaee and Farhi (2019b), all the results in the paper to arbitrary neoclassical production functions simply by replacing the input-output covariance operator with the

[3]When we refer to the price of producer-specific fixed factors, we rely on Lionel McKenzie's insight that any non-CRS production function $h(x)$ can be represented by a CRS technology $\tilde{h}(x,z) = zh(x/z)$ where z is a producer-specific fixed factor with supply $z = 1$. The marginal cost of $h(x)$ coincides with the marginal cost of $\tilde{h}(x,z)$, where the effect of scale in the former is captured by the (shadow) price of the fixed factor in the latter.

[4]Recall that primary factors $f \in \mathcal{F} \subset \mathcal{N}$ are captured as producer-specific fixed factors in factor markets with zero-returns-to-scale individual producers ($1-\varepsilon_f = 1$) aggregated linearly with no entry ($\widehat{\mathrm{d}\log\lambda}_{\pi,f} = 0$).

[5]The assumption that $\varepsilon_i, \gamma_i < 1$ ensures that entry is not socially wasteful. When it is violated, equilibrium reallocations affecting entry can reduce (but not increase) aggregate output to a first order.

input-output substitution operator instead. For a producer k with cost function $\mathbf{C}_k$, the Allen-Uzawa elasticity of substitution between inputs x and y is

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

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. We also use this definition for final demand aggregators.

The *input-output substitution operator* for producer k is defined as

$$\Phi_k(\Psi_{(i)},\Psi_{(j)}) = -\sum_{x,y\in N+F} \Omega_{kx}[\delta_{xy} + \Omega_{ky}(\theta_k(x,y)-1)]\Psi_{xi}\Psi_{yj}, \tag{59}$$

$$= \frac{1}{2}E_{\Omega^{(k)}}\Big((\theta_k(x,y)-1)(\Psi_i(x)-\Psi_i(y))(\Psi_j(x)-\Psi_j(y))\Big), \tag{60}$$

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.

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 when $\mathbf{C}_k$ has a CES form we recover the input-output covariance operator:

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

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.

All the results in the paper can be extended to general non-CES economies by simply replacing terms of the form $(\theta_k - 1)Cov_{\Omega^{(k)}}(\Psi_{(i)},\Psi_{(j)})$ by $\Phi_k(\Psi_{(i)},\Psi_{(j)})$.

Appendix E Details of Sectoral Models

For any sectoral model with heterogeneous firms in each sector, there is an isomorphic *companion* sectoral model with homogenous firms in each sector. The companion model assumes that all firms in a given sector $\mathcal{I}$ are identical with productivity shifter $A_{\mathcal{I}}$ and markup $\mu_{\mathcal{I}}$ defined by

$$A_{\mathcal{I}} = \frac{\mu_{\mathcal{I}}}{\overline{\mu}_{\mathcal{I}}} \left(\sum_{i \in \mathcal{I}} \overline{\lambda}_i^{\mathcal{I},\mathrm{B}} \left(\frac{A_i/\overline{A}_i}{\mu_i/\overline{\mu}_i} \right)^{\frac{\varepsilon_{\mathcal{I}}}{1-\varepsilon_{\mathcal{I}}}} \right)^{\frac{1-\varepsilon_{\mathcal{I}}}{\varepsilon_{\mathcal{I}}}} \qquad \text{and} \qquad \mu_{\mathcal{I}} = \frac{1}{\sum_{i \in \mathcal{I}} \lambda_i^{\mathcal{I},\mathrm{B}} \frac{1}{\mu_i}},$$

where for each $i \in \mathcal{I}$, we define $\lambda_{\mathcal{I}}^{\mathrm{B}} = \sum_{j \in \mathcal{I}} \lambda_j^{\mathrm{B}}$ and $\lambda_i^{\mathcal{I},\mathrm{B}} = \lambda_i^{\mathrm{B}}/\lambda_{\mathcal{I}}^{\mathrm{B}}$. Here we remind the reader that we use overlines to signal initial values when there is an ambiguity but we drop them when there is none. We denote by $\check{Y}$ the aggregate output in the companion model without heterogeneity within sectors.

If Y denotes aggregate output in a sectoral model with heterogeneity, we denote by $\check{Y}$ denote aggregate output in the companion model without heterogeneity.

Proposition 8 (Sectoral Aggregation)**.** *For any sectoral model with within-sector heterogeneity, the nonlinear response $\Delta \log Y$ of aggregate output to shocks to productivities and markups is equal to the nonlinear response $\Delta \log \check{Y}$ of aggregate output to shocks to sectoral productivities and markups in the companion model with no within-sector heterogeneity.*

Proof of Proposition 8. To prove this, for each industry with heterogeneous firms, we construct an isomorphic industry with homogeneous firms which has the same price, quantity and mass of entrants. To do this, consider some industry with heterogeneous firms, where we drop the industry subscript to cut down on notation. The equations that determine the industry's mass of entrants, prices and quantity produced are

$$\begin{aligned}
Y &= \left(\sum_i M_i y_i \right)^{1/\gamma} \\
y_i &= b_i q_i^{\varepsilon} \\
p_i^y &= \frac{\mu_i^y}{b_i} \frac{p_i^q}{\varepsilon} q_i^{1-\varepsilon} \\
p_i^q &= \frac{\mu_i^q p^{inputs}}{A_i} \\
P^Y &= \mu^Y \gamma_i p_i^y Y^{1-\frac{1}{\gamma}} \\
M_i &= b_i M, \\
M &= \frac{1}{\gamma} \mu^Y \sum_i \left(1 - \frac{\varepsilon}{\mu_i^q \mu_i^y} \right) \lambda_i,
\end{aligned}$$

where b_i are the exogenous taste/productivity shifters for each firm. The comparison industry with homogeneous firms is

$$
\begin{aligned}
Y_* &= (M_* y_*)^{\frac{1}{\gamma}} \\
y_* &= q_*^{\varepsilon} \\
p_*^y &= \mu_*^y \frac{p_*^q}{\varepsilon} q_*^{1-\varepsilon} \\
p_*^q &= \frac{\mu_*^q p_*^{inputs}}{A_*} \\
P_*^Y &= \mu_*^Y p_*^y Y^{1-1/\gamma} \\
M_* &= \frac{1}{\gamma} \mu_*^Y \left(1 - \frac{\varepsilon}{\mu_*^q \mu_*^y}\right) \lambda_*.
\end{aligned}
$$

We want to have μ^q, μ^y, μ^Y, A, such that we match the quantity $Y = Y_*$ and the price $P = P^*$ in the two cases. We need also want the mass of entrants to be the same.

$$M_* = M \tag{61}$$

hence

$$
\begin{aligned}
\mu_*^Y \left(1 - \frac{\varepsilon}{\mu_*^q \mu_*^y}\right) &= \mu^Y \sum_i \left(1 - \frac{\varepsilon}{\mu_i^q \mu_i^y}\right) \frac{\lambda_i}{\lambda_*} \\
&= \mu^Y \left(1 - (\varepsilon) \sum_i \frac{\delta_i}{\mu_i^q \mu_i^y}\right),
\end{aligned}
$$

where δ_i is firm i's sales shares in the industry. So, set

$$\mu_*^Y = \mu^Y \tag{62}$$

$$\mu_*^q \mu_*^y = \left(\sum_i \frac{\delta_i}{\mu_i^q \mu_i^y}\right)^{-1}. \tag{63}$$

To ensure that

$$P^Y = P_*^Y, \tag{64}$$

we need

$$\mu_*^y \mu_*^q \frac{1}{A_*} q_*^{1-\varepsilon} = \frac{\mu_i^y \mu_i^q}{b_i} \frac{1}{A_i} q_i^{1-\varepsilon} \tag{65}$$

and we know that

$$P^Y = \mu^Y \frac{\mu_i^y}{b_i} \frac{\mu_i^q p^{inputs}}{(\varepsilon)A_i} q_i^{1-\varepsilon} \gamma Y^{1-1/\gamma} \tag{66}$$

Hence

$$\left(\frac{b_i P^Y (\varepsilon) A_i}{Y^{1-1/\gamma} \gamma \mu^Y \mu_i^y \mu_i^q p^{inputs}} \right) = q_i^{1-\varepsilon} \tag{67}$$

Therefore,

$$\mu_*^y \mu_*^q \frac{1}{A_*} q_*^{1-\varepsilon} = \mu_i^y \mu_i^q \frac{1}{A_i} \left(\frac{P^Y(\varepsilon) A_i}{Y^{1-1/\gamma} \gamma \mu^Y \mu_i^y \mu_i^q p^{inputs}} \right)$$

$$q_* = \left(\left(\frac{P^Y(\varepsilon)}{Y_*^{1-1/\gamma} \gamma \mu^Y p^{inputs}} \right) \frac{A_*}{\mu_*^y \mu_*^q} \right)^{\frac{1}{1-\varepsilon}}$$

But we also must have

$$Y = \left(\sum_i M_i q_i^\varepsilon \right) = (M_* q_*^\varepsilon) = Y_* \tag{68}$$

In other words

$$M_* \left(\frac{P^Y(\varepsilon)}{Y_*^{1-1/\gamma} \gamma \mu^Y p^{inputs}} \frac{A_*}{\mu_*^y \mu_*^q} \right)^{\frac{\varepsilon}{1-\varepsilon}} = \left(\sum_i b_i M \left(\frac{P^Y(\varepsilon) A_i}{Y^{1-1/\gamma} \gamma \mu^Y \mu_i^y \mu_i^q p^{inputs}} \right)^{\frac{\varepsilon}{1-\varepsilon}} \right) \tag{69}$$

$$\left(\frac{A_*}{\mu_*^y \mu_*^q} \right)^{\frac{\varepsilon}{1-\varepsilon}} = \left(\sum_i b_i \left(\frac{A_i}{\mu_i^y \mu_i^q} \right)^{\frac{\varepsilon}{1-\varepsilon}} \right) \tag{70}$$

or

$$A_* = \mu_*^y \mu_*^q \left(\sum_i b_i \left(\frac{A_i}{\mu_i^y \mu_i^q} \right)^{\frac{\varepsilon}{1-\varepsilon}} \right)^{\frac{1-\varepsilon}{\varepsilon}} .$$

$$= \left(\sum_i b_i \left(\frac{A_i}{\mu_i^y \mu_i^q / (\mu_*^y \mu_*^q)} \right)^{\frac{\varepsilon}{1-\varepsilon}} \right)^{\frac{1-\varepsilon}{\varepsilon}} .$$

Let sectoral productivity be given by A_* and sectoral markups be given by μ_* where recall that $\mu_i^q \mu_i^y = \mu_i$. ■

The outer-elasticity $\gamma_{\mathcal{I}}$, which distinguishes models with IRS from those with DRS, is not relevant to how we aggregate firms within the sector since neither $A_{\mathcal{I}}$ nor $\mu_{\mathcal{I}}$ depend on $\gamma_{\mathcal{I}}$.

To break this problem into a within-sector and cross-sector problem, in vector notation, write

$$\mathrm{d}\log Y = \mathrm{d}\log \check{Y} = \sum_{\mathcal{I}} \frac{\mathrm{d}\log \check{Y}}{\mathrm{d}\log A_{\mathcal{I}}} \mathrm{d}\log A_{\mathcal{I}} + \sum_{\mathcal{I}} \frac{\mathrm{d}\log \check{Y}}{\mathrm{d}\log(\mu_{\mathcal{I}}, \mu^{Y}_{\mathcal{I}})} \mathrm{d}\log(\mu_{\mathcal{I}}, \mu^{Y}_{\mathcal{I}}).$$

We now differentiate a second time and evaluate the second derivative at the efficient marginal-pricing equilibrium. We use the fact that at the efficient point $\mathrm{d}\log A_{\mathcal{I}} = 0$ and, from the envelope theorem, that $\mathrm{d}\log \check{Y}/\,\mathrm{d}\log \mu_{\mathcal{I}} = 0$, we get a simpler expression for the loss function $\mathcal{L} = -(1/2)d^2 \log \check{Y}$ using

$$d^2 \log \check{Y} = \sum_{\mathcal{I}} \frac{\mathrm{d}\log \check{Y}}{\mathrm{d}\log A_{\mathcal{I}}} d^2 \log A_{\mathcal{I}} + \sum_{\mathcal{I},\mathcal{J}} \mathrm{d}\log(\mu_{\mathcal{I}}, \mu^{Y}_{\mathcal{I}})' \frac{d^2 \log \check{Y}}{\mathrm{d}\log(\mu_{\mathcal{I}}, \mu^{Y}_{\mathcal{I}})\,\mathrm{d}\log(\mu_{\mathcal{J}}, \mu^{Y}_{\mathcal{J}})} \mathrm{d}\log(\mu_{\mathcal{J}}, \mu^{Y}_{\mathcal{J}}), \quad (71)$$

where $\mathrm{d}\log \check{Y}/\,\mathrm{d}\log A_{\mathcal{I}} = \lambda^{B}_{\mathcal{I}}(\varepsilon_{\mathcal{I}})$ by Theorem 2. This expression can then be combined with the following lemma.

Using Lemma 6 below, it becomes apparent that: the first term in the loss function captures misallocation arising from distortions in relative producer sizes driven by the *dispersion* of markups/wedges within sectors; the second term captures misallocation arising from distortions in entry within sectors and relatives sizes across sectors arising driven by the *levels* of markups. The losses increase with the returns to scale and go to infinity in the constant-returns limit where $1 - \varepsilon_{\mathcal{I}}$ goes to zero.

Lemma 6. *At the efficient marginal-cost pricing equilibrium, changes in sectoral markups and productivities in the companion model are related to changes in markups/wedges in the original model according to*

$$\mathrm{d}\log \mu_{\mathcal{I}} = \mathbb{E}_{\lambda^{\mathcal{I},B}}(\mathrm{d}\log \mu), \quad \mathrm{d}\log A_{\mathcal{I}} = 0, \quad \textit{and} \quad d^2 \log A_{\mathcal{I}} = \frac{1}{1-\varepsilon_{\mathcal{I}}} Var_{\lambda^{\mathcal{I},B}}(\mathrm{d}\log \mu),$$

where these expressions denote within-sector weighted expectations and variances of the changes in markups/wedges $\mathrm{d}\log \mu_i$ *in the original model, with weights given by the within-sectoral sales share distribution* $\lambda^{\mathcal{I},B}_i$.

Proof of Lemma 6. First, we solve out for A as a function of primitives. Using the same notation as in the proof of Proposition 8, note that

$$\lambda_i = \frac{M p^{y}_i y_i}{P^Y Y}.$$

Use the fact that

$$p_i^y = \frac{1}{\varepsilon}\mu_i^q \mu_i^y \left(\frac{q_i}{b_i}\right)^{1-\varepsilon} p^{inputs} = P^Y. \tag{72}$$

Hence

$$y_i = b_i^{1-\varepsilon} q_i^{\varepsilon} = b_i^{1-\varepsilon}\left(\frac{P^Y(\varepsilon) b_i^{1-\varepsilon}}{\mu_i^q \mu_i^y p^{inputs}}\right)^{\frac{\varepsilon}{1-\varepsilon}}. \tag{73}$$

Next note that, firm i's market share δ_i is given by

$$\delta_i = \frac{My_i}{Y} = \frac{Mb_i^{1-\varepsilon}\left(\frac{P^Y(\varepsilon)b_i^{1-\varepsilon}}{\mu_i^q\mu_i^y p^{inputs}}\right)^{\frac{\varepsilon}{1-\varepsilon}}}{\sum_j Mb_j^{1-\varepsilon}\left(\frac{P^Y(\varepsilon)b_j^{1-\varepsilon}}{\mu_j^q\mu_j^y p^{inputs}}\right)^{\frac{\varepsilon}{1-\varepsilon}}}$$
$$= \frac{b_i\mu_i^{\frac{-\varepsilon}{1-\varepsilon}}}{\sum_j b_j\mu_j^{\frac{-\varepsilon}{1-\varepsilon}}}.$$

Hence, substituting in, we have

$$\mu_* = \left(\sum_i \frac{\delta_i}{\mu_i}\right)^{-1},$$
$$= \left(\sum_i \frac{b_i\mu_i^{-\frac{1}{1-\varepsilon}}}{\sum_j b_j\mu_j^{\frac{-\varepsilon}{1-\varepsilon}}}\right)^{-1}$$

which means we can write

$$A = \left(\sum_i \frac{b_i\mu_i^{-\frac{1}{1-\varepsilon}}}{\sum_j b_j\mu_j^{\frac{1-\varepsilon-1}{1-\varepsilon}}}\right)^{-1}\left(\sum_i b_i\left(\mu_i\right)^{\frac{1-\varepsilon-1}{1-\varepsilon}}\right)^{\frac{1-\varepsilon}{\varepsilon}}. \tag{74}$$

First consider the derivatives of the sectoral productivity shifter

$$\log A = -\frac{1}{\left(\sum_i b_i\left(\mu_i^{-\frac{1}{1-\varepsilon}}\right)\right)}\left(\sum_i b_i\mu_i^{-\frac{1}{1-\varepsilon}}\left(-\frac{1}{1-\varepsilon}\right)d\log\mu_i\right) + \frac{\frac{1}{1-\varepsilon}}{\frac{1}{1-\varepsilon}-1}\frac{1-\frac{1}{1-\varepsilon}}{\left(\sum_i b_i\left(\mu_i\right)^{1-\frac{1}{1-\varepsilon}}\right)}\left(\sum_i b_i\mu_i^{1-\frac{1}{1-\varepsilon}}d\log\mu_i\right)$$
$$= \frac{\frac{1}{1-\varepsilon}}{\left(\sum_i b_i\left(\mu_i^{-\frac{1}{1-\varepsilon}}\right)\right)}\left(\sum_i b_i\mu_i^{-\frac{1}{1-\varepsilon}}d\log\mu_i\right) - \frac{1}{1-\varepsilon}\frac{1}{\left(\sum_i b_i\left(\mu_i\right)^{1-\frac{1}{1-\varepsilon}}\right)}\left(\sum_i b_i\mu_i^{1-\frac{1}{1-\varepsilon}}d\log\mu_i\right)$$
$$d^2\log A = -\frac{\frac{1}{1-\varepsilon}^2}{\left(\sum_i b_i\left(\mu_i^{-\frac{1}{1-\varepsilon}}\right)\right)}\left(\sum_i b_i\mu_i^{-\frac{1}{1-\varepsilon}}d\log\mu_i^2\right) + \frac{\frac{1}{1-\varepsilon}^2}{\left(\sum_i b_i\mu_i^{-\frac{1}{1-\varepsilon}}\right)}\left(\sum_i b_i\mu_i^{-\frac{1}{1-\varepsilon}}d\log\mu_i\right)^2$$

$$+\frac{1}{1-\varepsilon}\frac{(1-\frac{1}{1-\varepsilon})}{\left(\sum_i b_i(\mu_i)^{1-\frac{1}{1-\varepsilon}}\right)}\left(\sum_i b_i\mu_i^{1-\frac{1}{1-\varepsilon}}d\log\mu_i\right)^2-\frac{1}{1-\varepsilon}\frac{(1-\frac{1}{1-\varepsilon})}{\left(\sum_i b_i(\mu_i)^{1-\frac{1}{1-\varepsilon}}\right)}\left(\sum_i b_i\mu_i^{1-\frac{1}{1-\varepsilon}}d\log\mu_i^2\right)$$
$$=-\frac{1}{1-\varepsilon}\left(\sum_i b_i d\log\mu_i^2\right)+\frac{1}{1-\varepsilon}\left(\sum_i b_i d\log\mu_i\right)^2.$$

Obviously, at the efficient point $d\log A=0$.

Now consider the log-derivative of the sectoral markup

$$d\log\mu_*=-\frac{1}{\mu_*}\left(-\sum_i\frac{1}{1-\varepsilon}\frac{b_i\mu_i^{-\frac{1}{1-\varepsilon}}d\log\mu_i}{\sum_j b_j\mu_j^{\frac{1-\varepsilon-1}{1-\varepsilon}}}-\frac{1-\varepsilon-1}{1-\varepsilon}\frac{\sum_i b_i\mu_i^{-\frac{1}{1-\varepsilon}}}{\left(\sum_j b_j\mu_j^{\frac{1-\varepsilon-1}{1-\varepsilon}}\right)^2}\sum_j b_j d\log\mu_j\right)$$
$$=-\left(-\frac{1}{1-\varepsilon}\sum b_i d\log\mu_i-\frac{1-\varepsilon-1}{1-\varepsilon}\sum_j b_j d\log\mu_j\right)$$
$$=\sum_j b_j d\log\mu_j.$$

■

Appendix F Relaxing Homotheticity/Iso-elasticity

In this section, we relax Assumption 1 by considering how the model changes if (i) entry happens via a non-iso-elastic Kimball (1995) aggregator, and (ii) if the extent of decreasing returns to scale is variable.

F.1 Relaxing IRS

For the IRS benchmark, we can relax the assumption that entry happens via a CES aggregator by using the Kimball demand system instead. In other words, index firms in market i by some parameter θ, and suppose the production function is given by

$$y_i(\theta)=A_i(\theta)\left[f_i(x_{ij}(\theta))\right]^{\varepsilon_i},\tag{75}$$

where f_i has constant returns to scale. Next, suppose that the inputs into the production function are defined implictly via the equation:

$$1=\int\Upsilon_{ij}\left(\frac{x_{ij}(\theta,\theta')}{x_{ij}(\theta)}\right)M_j(\theta')d\theta',\tag{76}$$

where Υ_{ij} is an increasing concave function and $M_j(\theta)$ is the mass of type θ firms in $j \in \mathcal{N}$. The resource constraint for the output of this firm is then

$$y_i(\theta) = \sum_j \int x_{ji}(\theta', \theta) M_j(\theta') d\theta' + c_i(\theta'). \tag{77}$$

Let $P(i, j)$ be the marginal cost of input $x_{ij}(\theta)$. Because of homotheticity, we can consider the marginal cost of $x_{ij}(\theta)$ as depending only on $\{p_j(\theta'), M_j(\theta')\}_{\theta'}$. Define for each $(i, j) \in \mathcal{N}^2$, the linear operator $s(i, j) : L_2(\mathbb{R}) \to \mathbb{R}$

$$s(i, j) \cdot z = \int \left(\frac{p_i(\theta') x_{ij}(\theta, \theta') M_i(\theta')}{P(i, j) x_{ij}(\theta)} \right) z(\theta') d\theta. \tag{78}$$

Then we can write the change in the marginal cost of x_{ij}

$$d \log P(i, j) = s(i, j) \cdot d \log p_j - s(i, j) \cdot \left[(\delta_{ij} - 1) d \log M_j \right], \tag{79}$$

where

$$\delta_{ij}(\theta) = \left(\int \Upsilon' \left(\frac{x_{ij}(\theta, \theta')}{x_{ij}(\theta)} \right) \frac{x_{ij}(\theta, \theta')}{x_{ij}(\theta)} M_j(\theta') d\theta' \right)^{-1}. \tag{80}$$

By homotheticity, $\delta_{ij}(\theta)$ is not a function of θ. The variable $\delta_{ij} > 1$ measures the love-of-variety effect in this model.

The Proposition below generalizes Theorem 3 to an economy with Kimball demand.

Proposition 9. *The response of aggregate output to shocks* $(\mathrm{d} \log A, \mathrm{d} \log \mu)$ *is given by*

$$d \log Y = \sum_{ij} \lambda^F_{ij} \left(s(i, j) \cdot (\varepsilon_i) \left(d \log \mu_i(\theta) - A_i(\theta) \right) - (1 - \varepsilon_i) s(i, j) \cdot \left(d \log \lambda_i(\theta) - d \log \hat{\lambda}_\pi(\theta) \right) \right)$$
$$+ \sum_{ij} \lambda^F_{ij} \left(s(i, j) \cdot (\delta_{ij} - 1) d \log \hat{\lambda}_\pi(\theta) \right).$$

where λ^F_{ij} *is the forward Domar weight of the price of the Kimball aggregator associated with i's inputs from j and* $s(i, j)$ *is the sales distribution of varieties in j who sell to i.*

This equation has a very similar form to Theorem 3 with similar intuition. Below, we derive Proposition 9, and also generalize the Forward and Backward propagation equations in Propositions 1 and 2.

By Shephard's lemma

$$d \log p_i(\theta) = -d \log A_i(\theta) + d \log \mu_i(\theta) + \sum_j \Omega^F_{ij} d \log P(i, j) + \frac{1 - \varepsilon_i}{\varepsilon_i} d \log y_i(\theta)$$

$$= -d\log A_i(\theta) + d\log \mu_i(\theta) + \sum_j \Omega^F_{ij} d\log P(i,j) + \frac{1-\varepsilon_i}{\varepsilon_i} d\log \lambda_i(\theta)$$
$$- \frac{1-\varepsilon_i}{\varepsilon_i} d\log M_i(\theta) - \frac{1-\varepsilon_i}{\varepsilon_i} d\log p_i(\theta)$$
$$= -(\varepsilon_i) d\log A_i(\theta) + (\varepsilon_i) d\log \mu_i(\theta) + (\varepsilon_i)\sum_j \Omega^F_{ij} d\log P(i,j)$$
$$+ (1-\varepsilon_i)\left(d\log \lambda_i(\theta) - d\log M_i(\theta)\right)$$

Therefore

$$d\log P(i,j) = s(i,j)\cdot\left(-(\varepsilon_i) d\log A_i(\theta) + (\varepsilon_i) d\log \mu_i(\theta) + (\varepsilon_i)\sum_j \Omega^F_{ij} d\log P(i,j)\right)$$
$$+ s(i,j)\cdot\left((1-\varepsilon_i)\left(d\log \lambda_i(\theta) - d\log M_i(\theta)\right)\right) - s(i,j)\cdot\left[(\delta_{ij}-1) d\log M_j\right]$$
$$= \left(s(i,j)\cdot(\varepsilon_i) d\log \frac{\mu_i(\theta)}{A_i(\theta)} + (\varepsilon_i)\sum_j \Omega^F_{ij} d\log P(i,j)\right)$$
$$+ (1-\varepsilon_i) s(i,j)\cdot\left(d\log \lambda_i(\theta) - d\log M_i(\theta)\right) - s(i,j)\cdot\left[(\delta_{ij}-1) d\log M_j\right]$$

We also have that

$$\lambda_{\pi_i}(\theta) = \left(1 - \frac{\varepsilon_i}{\mu_i(\theta)}\right)\lambda_i(\theta). \tag{81}$$

Define the function $\zeta_j(i,\theta)$ to be the mass of entrant j mapped to (i,θ). Zero-profit condition for type j entrant is

$$E_{\zeta_j}\left(\lambda_{\pi_i}(\theta)\right) = P_E, \tag{82}$$

where the expectation is with respect to ζ_j. We also have

$$M_i(\theta) = \int_E \zeta_j(i,\theta) M_{E,j}\, dj. \tag{83}$$

So we can write

$$d\log \lambda_{\pi_i}(\theta) = d\log \lambda_i(\theta) + \frac{\frac{\varepsilon_i}{\mu_i(\theta)}}{\left(1 - \frac{\varepsilon_i}{\mu_i(\theta)}\right)} d\log \mu_i(\theta) \tag{84}$$

$$\sum_i \int \zeta_j(i,\theta) \frac{\lambda_{\pi_i}(\theta)}{M_i(\theta)}\left(d\log \lambda_{\pi_i}(\theta) - d\log M_i(\theta)\right) d\theta = P_E d\log P_E \tag{85}$$

$$d\log M_i(\theta) = \frac{\int_E \zeta_j(i,\theta) M_{E,j} d\log M_{E,j} dj}{\int_E \zeta_j(i,\theta) M_{E,j} dj}. \tag{86}$$

Let $\tilde{\zeta} : E \to \mathbb{R}^{\mathcal{N}}$and $\lambda_\pi : \mathbb{R}^{\mathcal{N}} \to \mathbb{R}^{\mathcal{N}}$be linear operators. Then we can write

$$d\log M_i(\theta) = \tilde{\zeta} \cdot d\log M_E \tag{87}$$

$$\tilde{\zeta}^* \cdot \lambda_\pi \cdot d\log\lambda_\pi - \tilde{\zeta}^* \cdot \lambda_\pi \cdot d\log M = P_E d\log P_E \tag{88}$$

$$\begin{aligned}
\tilde{\zeta}^* \cdot \lambda_\pi \cdot d\log\lambda_\pi - \tilde{\zeta}^* \cdot \lambda_\pi \cdot \tilde{\zeta} \cdot d\log M_E &= P_E d\log P_E \\
\left(\tilde{\zeta}^* \cdot \lambda_\pi \cdot \tilde{\zeta}\right)^{-1} \left(\tilde{\zeta}^* \cdot \lambda_\pi \cdot d\log\lambda_\pi - P_E d\log P_E\right) &= d\log M_E \\
\tilde{\zeta} \cdot \left(\tilde{\zeta}^* \cdot \lambda_\pi \cdot \tilde{\zeta}\right)^{-1} \left(\tilde{\zeta}^* \cdot \lambda_\pi \cdot d\log\lambda_\pi - P_E d\log P_E\right) &= d\log M,
\end{aligned}$$

where $\tilde{\zeta}^*$is the adjoint operator. Define

$$d\log\hat{\lambda}_\pi = \tilde{\zeta} \cdot \left(\tilde{\zeta}^* \cdot \lambda_\pi \cdot \tilde{\zeta}\right)^{-1} \tilde{\zeta}^* \cdot \lambda_\pi \cdot d\log\lambda_\pi$$

Hence, the forward equation becomes

$$\begin{aligned}
d\log P(i,j) &= \left(s(i,j) \cdot (\varepsilon_i) d\log\frac{\mu_i(\theta)}{A_i(\theta)} + (\varepsilon_i)\sum_j \Omega_{ij} d\log P(i,j) \right) && (89)\\
&+ \left((1-\varepsilon_i)s(i,j) \cdot (d\log\lambda_i(\theta) - d\log M_i(\theta))\right) - s(i,j) \cdot \left[(\delta_{ij}-1) d\log M_j\right] && (90)\\
&= \left(s(i,j) \cdot (\varepsilon_i) d\log\frac{\mu_i(\theta)}{A_i(\theta)} + (\varepsilon_i)\sum_j \Omega_{ij} d\log P(i,j) \right) && (91)\\
&+ \left((1-\varepsilon_i)s(i,j) \cdot \left(d\log\lambda_i(\theta) - d\log\lambda_\pi(\theta) + \tilde{\zeta} \cdot \left(\tilde{\zeta}^* \cdot \lambda_\pi \cdot \tilde{\zeta}\right)^{-1} \tilde{\zeta}^* \cdot \lambda_\pi \cdot \lambda_E \cdot \Omega^E d\log P(i,j)\right)\right) \\
&- s(i,j) \cdot (\delta_{ij}-1) d\log\lambda_\pi(\theta) - \left[s(i,j) \cdot (\delta_{ij}-1)\right] \tilde{\zeta} \cdot \left(\tilde{\zeta}^* \cdot \lambda_\pi \cdot \tilde{\zeta}\right)^{-1} \tilde{\zeta}^* \cdot \lambda_\pi \cdot \lambda_E \cdot \Omega^E d\log P(i,j) && (92)
\end{aligned}$$

This is a linear system in $d\log P(i,j)$. Group (i,j) together and write this linear system as a $\mathcal{N}^2 \times 1$ vector, with an appropriately defined Ψ^F,then we have

$$d\log P(l,m) = \sum_{ij} \Psi^F(lm,ij) \left(s(i,j) \cdot (\varepsilon_i) d\log\frac{\mu_i(\theta)}{A_i(\theta)} + (1-\varepsilon_i)s(i,j) \cdot \left(d\log\lambda_i(\theta) - d\log\hat{\lambda}_\pi(\theta)\right)\right)$$

$$-\sum_{ij} \Psi^F(lm, ij)\Big(s(i,j)\cdot(\delta_{ij}-1)d\log\hat{\lambda}_{\pi}(\theta)\Big).$$

This is the generalization to Theorem 3 and Proposition 1, showing that those results survive generalization.

Next, to pin down $d\log\lambda$, we need an analogue to the backward equations.

$$y_i(\theta) = \sum_j \int x_{ji}(\theta',\theta)M_j(\theta')M_j(\theta')d\theta' + c_i(\theta'). \tag{93}$$

$$\begin{aligned}\lambda_i(\theta) &= M_i(\theta)p_i(\theta)y_i(\theta)\\ &= M_i(\theta)p_i(\theta)\sum_j \int x_{ji}(\theta',\theta)M_j(\theta')d\theta'.\end{aligned}$$

Define

$$\sigma_{ji} = -\frac{\Upsilon'\left(\frac{y_{ji}(\theta)}{y_{ji}}\right)}{-\frac{y_{ji}(\theta)}{y_{ji}}\Upsilon''\left(\frac{y_{ji}(\theta)}{y_{ji}}\right)}, \tag{94}$$

where $y_{ji}(\theta) = \int x_{ji}(\theta',\theta)M_j(\theta')d\theta'$ and y_{ji} is defined implicitly via $1 = \int \Upsilon_{ji}\left(\frac{y_{ji}(\theta)}{y_{ji}}\right)M_j(\theta)d\theta$. Intuitively, because of homotheticity, we can assume that an intermediary purchases y_{ji} and sells it at marginal cost to all θ types in industry j. The quantity purchased by the intermediary from firm θ' in industry i is $y_{ji}(\theta')$ and the total output of the intermediary is y_{ji}.

The variable σ_{ji} is the price-elasticity of residual demand.

$$\begin{aligned}-d\log\delta_{ij} &= \frac{\int\left(\Upsilon'\left(\frac{y_{ij}(\theta)}{y_{ij}}\right)\frac{y_{ij}(\theta)}{y_{ij}}M_j(\theta)\right)\left[\frac{\frac{y_{ij}(\theta)}{y_{ij}}\Upsilon''\left(\frac{y_{ij}(\theta)}{y_{ij}}\right)}{\Upsilon'\left(\frac{y_{ij}(\theta)}{y_{ij}}\right)}d\log\left(\frac{y_{ij}(\theta)}{y_{ij}}\right)+d\log\frac{y_{ij}(\theta)}{y_{ij}}+d\log M_j(\theta)\right]d\theta}{\int\Upsilon'\left(\frac{y_{ij}(\theta)}{y_{ij}}\right)\frac{y_{ij}(\theta)}{y_{ij}}M_j(\theta)d\theta.}\\
&= \frac{\int\left(\frac{p_j(\theta)}{\delta_{ij}P(i,j)}\frac{y_{ij}(\theta)}{y_{ij}}M_j(\theta)\right)\left[\left(1-\frac{1}{\sigma_{ji}(\theta)}\right)d\log\left(\frac{y_{ij}(\theta)}{y_{ij}}\right)+d\log M_j(\theta)\right]d\theta}{\int\frac{p_j(\theta)}{\delta_{ij}P(i,j)}\frac{y_{ij}(\theta)}{y_{ij}}M_j(\theta)d\theta.}\\
&= s(i,j)\cdot\left[\left(1-\frac{1}{\sigma_{ji}(\theta)}\right)d\log\left(\frac{y_{ij}(\theta)}{y_{ij}}\right)+d\log M_j(\theta)\right]\end{aligned} \tag{95}$$

$$d\log\left(\frac{y_{ij}(\theta)}{y_{ij}}\right) = d\log(\Upsilon')^{-1}_{ji}\left(\frac{p_i(\theta)}{\delta_{ji}P(j,i)}\right) \tag{96}$$

Hence

$$d\log\left(\frac{y_{ij}(\theta)}{y_{ij}}\right) = \sigma_{ij}(\theta)\left(d\log p_i(\theta) - d\log\delta_{ji} - d\log P(j,i)\right). \tag{97}$$

Use this in

$$\begin{aligned}\lambda_i(\theta) &= M_i(\theta)p_i(\theta)y_i(\theta)\\ &= M_i(\theta)p_i(\theta)\sum_j y_{ji}(\Upsilon')^{-1}_{ij}\left(\frac{p_i(\theta)}{\delta_{ji}P(j,i)}\right),\end{aligned}$$

where the final line follows from homotheticity. Hence

$$\begin{aligned}d\log\lambda_i(\theta) = d\log M_i(\theta) + d\log p_i(\theta) + \sum_j \frac{y_{ji}(\theta)}{y_{ji}}\left(\sigma_{ji}(\theta)\left(d\log p_i(\theta) - d\log\delta_{ji} - d\log P(j,i)\right)\right)\\ + \sum_j \frac{y_{ji}(\theta)}{y_{ji}} d\log y_{ji}.\end{aligned} \tag{98}$$

Next, use

$$y_{ji} = \frac{\varepsilon_j}{\bar{\mu}_j}\frac{\Omega^F_{ji}\lambda_j}{P(j,i)} \tag{99}$$

coupled with

$$\Omega^F_{ji} d\log\Omega^F_{ji} = (1-\theta_j)Cov_j(d\log P(j,m), I_{(i)}) \tag{100}$$

to get

$$d\log y_{ji} = -d\log\bar{\mu}_j + (1-\theta_j)Cov_j(d\log P(j,m), I_{(i)}) + d\log\lambda_j - d\log P(j,i), \tag{101}$$

where

$$\bar{\mu}_j = \left(\int \frac{\lambda_i(\theta)M_i(\theta)}{\int\lambda_i(\theta)M_i(\theta)d\theta}\frac{1}{\mu_i(\theta)}d\theta\right)^{-1}, \tag{102}$$

so

$$-d\log\bar{\mu}_i = \int \frac{\lambda_i(\theta)M_i(\theta)}{\int\lambda_i(\theta)M_i(\theta)d\theta}\frac{1}{\mu_i(\theta)}\left[-d\log\mu_i(\theta) + d\log\lambda_i(\theta) + d\log M_i(\theta) - d\log\lambda_i\right]d\theta \tag{103}$$

Finally, use the fact that

$$\lambda_i = \sum_j \frac{\varepsilon_j}{\bar{\mu}_j}\Omega^F_{ji}\lambda_j + \sum_j \Omega^E_{ji}\lambda_{E,j} \tag{104}$$

to get

$$d\lambda_i = \sum_j d\lambda_j \frac{\varepsilon_j}{\bar{\mu}_j}\Omega^F_{ji} - \sum_j \frac{\lambda_j}{\bar{\mu}_j}(\varepsilon_j)\Omega^F_{ji} d\log\bar{\mu}_j + \sum_j \frac{\lambda_j}{\bar{\mu}_j}(\varepsilon_j)\Omega^F_{ji} d\log\Omega^F_{ji} + \sum_j d\lambda_{E,j}\Omega^E_{ji}. \tag{105}$$

Equations (92), (95), (98), (100), (101), (103), (105) jointly complete the characterization.

F.2 Relaxing DRS

Suppose that

$$Y = M_i A_i f_i\left(\left\{x_{ij}\right\}_j\right), \tag{106}$$

where we do not impose homotheticity on f_i. This means that every sector is DRS, but need not be homothetic.

Proposition 10. *The response of aggregate output to shocks* $(\mathrm{d}\log A, \mathrm{d}\log\mu, \mathrm{d}\log\mu^Y)$ *is given by*

$$d\log Y = -\lambda^{F'}\left(d\log\mu^Y - d\log A + (\varepsilon)d\log\mu\right) - \lambda^{F'}(1-\varepsilon)\left(d\log\lambda + d\log(1-\varepsilon) - d\log\hat{\lambda}_\pi\right),$$

where $\mathrm{d}\log(1-\varepsilon)$ *is the change in the returns to scale in each market.*

Proposition 10 generalizes Proposition 3. Below, we derive the Forward and Backward propagation equations in 1 and 2.

Define

$$\lambda^Y = PY = \mu^Y\lambda^y$$
$$\lambda^y = pyM$$
$$\lambda_\pi = \frac{1}{\mu^Y}\left(1 - \frac{\varepsilon}{\mu}\right)\lambda^Y$$

which implies that

$$d\log\lambda_\pi = -d\log\mu^Y - \frac{\frac{\varepsilon}{\mu}}{\left(1-\frac{\varepsilon}{\mu}\right)}\left[d\log(\varepsilon) - d\log\mu\right] + d\log\lambda$$
$$d\log M = d\log\hat{\lambda}_\pi - d\log\hat{P}$$

$$P = \mu^Y\frac{dC}{dY} = \mu^Y C/Y$$

$$d\log P = d\log \mu^Y + d\log p$$

$$\begin{aligned}
d\log p &= d\log\mu - d\log(\varepsilon) - d\log y_i + \Omega^F d\log P + \frac{\partial \log C_i}{\partial \log y_i}\left(d\log y_i - d\log A_i\right)\\
&= d\log\mu - d\log(\varepsilon) - d\log y_i + \Omega^F d\log P + \frac{1}{\varepsilon}\left(d\log y_i - d\log A_i\right)\\
&= d\log\mu - d\log(\varepsilon) + \Omega^F d\log P + \frac{1-\varepsilon}{\varepsilon}\left(d\log\lambda - d\log p - d\log M - \frac{1}{1-\varepsilon} d\log A_i\right)\\
d\log p &= (\varepsilon) d\log\mu - (\varepsilon) d\log(\varepsilon) + (\varepsilon)\Omega^F d\log P + (1-\varepsilon)\left(d\log\lambda - d\log M\right) - d\log A\\
d\log P &= d\log\mu^Y + d\log p\\
&= d\log\mu^Y + \left[(\varepsilon) d\log\mu + d(1-\varepsilon) + (\varepsilon)\Omega^F d\log P + (1-\varepsilon)\left(d\log\lambda - d\log M\right) - d\log A\right]\\
&= d\log\mu^Y + (\varepsilon) d\log\mu + d(1-\varepsilon) + (\varepsilon)\Omega^F d\log P\\
&+ (1-\varepsilon)\left(d\log\lambda - d\log\hat{\lambda}_\pi + d\log\hat{P}\right) - d\log A\\
(I-\Omega^F) d\log P &= d\log\mu^Y + (\varepsilon) d\log\mu + (1-\varepsilon)\left(d\log\lambda + d\log(1-\varepsilon) - d\log\hat{\lambda}_\pi\right) - d\log A
\end{aligned}$$

This last equation generalizes the forward equations in Proposition 1.

To get the backward equation, assuming some separability, we can write

$$f_i\left(\left\{x_{ij}\right\}_j\right) = f_i(q_i), \tag{107}$$

where q_i is CRS function of inputs. We can write

$$\begin{aligned}
\Omega_{ij} &= \frac{M_i P_j x_{ij}}{P_i Y_i} = \frac{P_j x_{ij}}{\mu^Y p_i y_i} = \frac{P_j x_{ij}}{\mu^Y \mu_i^y (\varepsilon_i) \mu_i^q mc_i q_i} = \frac{1}{\mu_i \mu_i^Y}\frac{1}{\varepsilon_i}\frac{p_j x_{ij}}{mc_i q_i}\\
d\log\Omega_{ij} &= -d\log\mu_i\mu_i^Y + d\log\gamma_i - d\log(\varepsilon_i) + d\log\left(\frac{p_j x_{ij}}{mc_i q_i}\right)
\end{aligned}$$

Denote the super-elasticity by $\frac{\partial^2 \log f_i}{\partial \log q_i^2} = \kappa_i$. Then we can write

$$d(\varepsilon_i) = \kappa_i\left(d\log\lambda_i^q - d\log p_i^q\right) = d\log\lambda_i^q - d\log\lambda_i^y. \tag{108}$$

Hence,

$$d(1-\varepsilon) = d\lambda_i^y - d\lambda_i^q \tag{109}$$

and

$$d\log\lambda_i^q = \frac{1}{\kappa_i - 1}\left(\kappa_i d\log p_i^q - d\log\lambda_i^y\right). \tag{110}$$

Hence

$$
\begin{aligned}
d\log\Omega_{ij} &= -d\log\mu_i\mu_i^Y - \kappa_i\left(d\log\lambda_i^q - d\log p_i^q\right) + d\log\left(\frac{p_j x_{ij}}{mc_i q_i}\right) \\
&= -d\log\mu_i\mu_i^Y - \frac{\kappa_i}{\varepsilon_i}\left(\frac{1}{\kappa_i - 1}\left(\kappa_i d\log p_i^q - d\log\lambda_i^y\right) - d\log p_i^q\right) + d\log\left(\frac{p_j x_{ij}}{mc_i q_i}\right) \\
&= -d\log\mu_i\mu_i^Y - \frac{\kappa_i}{\varepsilon_i}\left(\frac{1}{\kappa_i - 1}\left(\kappa_i d\log p_i^q - d\log\lambda_i^y\right) - d\log p_i^q\right) + (1-\theta_i)Cov_i(d\log P, I_{(i)}) \\
&= -d\log\mu_i\mu_i^Y - \frac{\kappa_i}{\varepsilon_i}\frac{1}{\kappa_i - 1}\left(d\log p_i^q - d\log\lambda_i^y\right) + (1-\theta_i)Cov_i(d\log P, I_{(i)}) \\
&= -d\log\mu_i\mu_i^Y - \frac{\kappa_i}{\varepsilon_i}\frac{1}{\kappa_i - 1}\left(\sum_j \Omega_{ij}P_j - d\log\lambda_i^y\right) + (1-\theta_i)Cov_i(d\log P, I_{(i)})
\end{aligned}
$$

Finally, combine this with

$$
d\lambda^{y'} = d\lambda^{y'}\Omega + \lambda^{y'}d\Omega + d\lambda_E\Omega^E \tag{111}
$$

to pin down the backward equations, which is the equivalent of Proposition 2.

Appendix G Mapping Model to Data

Our calibrated model is sectoral in the formal sense defined in the paper. Our calibration is very similar to Baqaee and Farhi (2019a), and we borrow much of the following discussion from the Appendix of that paper.

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.

G.1 Input-Output and Aggregate Data

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 backward expenditure share matrix Ω^B at the industry level. 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^B is negative, we set it to zero.

In Appendix H, we use alternative ways of estimating the markups. 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 forward matrix $\Omega^F = \mu\Omega$, also at the industry level. 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 sectoral structure along the lines of Section 7, 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 industry $\Omega^F_{ik} = \Omega^F_{jk}$. 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.

G.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.

G.2.1 Production Function Estimation Approach

This is our benchmark method for estimating markups, and the results in the main body of the paper use this approach. For reference, we will call this the production function estimation (or PF) markups.

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) (OP) 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 **?**, 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) (ACF). 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).

In our benchmark estimates, we treat SG&A as a fixed cost. However, for robustness, following De Loecker et al. (2019), we also compute markups using an approach where SG&A is treated as a variable input in production. We call these the PF2 markups. The overall estimation is still done via the ACF-corrected OP method (with CAPX as a proxy).

Finally, before feeding these markup estimates into the structural model, we winsorize the markups at the 20th and 80th percentile to reduce the influence of outliers.

G.2.2 User Cost Approach

Our second approach to measuring markups is the user-cost approach (UC) markups. The idea here is to recover the profits of a firm by subtracting total costs from revenues. To compute total cost, we must measure the 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,

$$\frac{OS_{i,t}}{K_{i,t}} = r_{k_i,t} + \left(1 - \frac{1}{\mu_i}\right) \frac{sales_{i,t}}{K_{i,t}}, \tag{112}$$

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 (112) for a time-varying firm-level measure of the markup. We winsorize markups at the 5-95th percentile by year.

G.2.3 Accounting Profits Approach

The final approach to estimating markups is the accounting profits approach (AC). 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). The advantage of this approach is its simplicity.

Appendix H Additional Quantitative Results

IRS, $\varepsilon = 0.875$	No Entry	Entry uses Factors	Entry uses Goods and Factors
PF2 Markups	12%	39%	47%
UC Markups	3.0%	23%	34%
AC Markups	4.5%	54%	75%
IRS, $\varepsilon = 0.75$	No Entry	Entry uses Factors	Entry uses Goods and Factors
PF2 Markups	24%	32%	31%
UC Markups	7.2%	15%	17%
AC Markups	11%	14%	14%
DRS, $\varepsilon = 0.875$			
PF2 Markups	19%	25%	25%
UC Markups	6.0%	11%	11%
AC Markups	8.2%	13%	12%
DRS, $\varepsilon = 0.75$			
PF2 Markups	9.0%	28%	29%
UC Markups	4.8%	40%	43%
AC Markups	2.6%	18%	19%

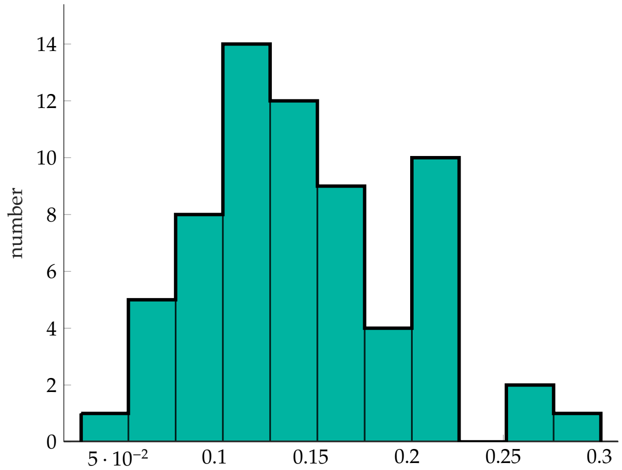
Table 2: The gains from moving to the efficient allocation. The IRS specification sets $\gamma_{\mathcal{I}} = \varepsilon_{\mathcal{I}} = 0.875$ and uses an imperfect-substitutes interpretation. The DRS specification sets $\gamma_{\mathcal{I}} = 1$, $\varepsilon_{\mathcal{I}} = 0.875$ and uses a perfect-substitutes interpretation.

Appendix I Second-Best Policy

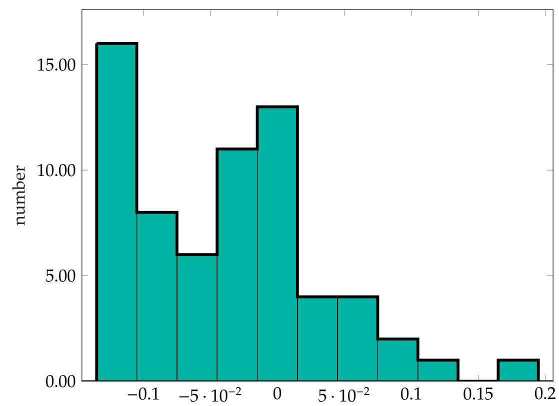
I.1 Bang for Buck of Marginal Policy Interventions

We end this section by considering the effect of a marginal policy intervention in the decentralized equilibrium. Figure 6 shows the bang-for-buck elasticity of aggregate output with respect to a marginal entry subsidy (a form of industrial policy) or markup reduction (a form of competition policy) in different industries. The elasticity is scaled by the revenues associated with the intervention, as in Section I, to make the magnitudes comparable.

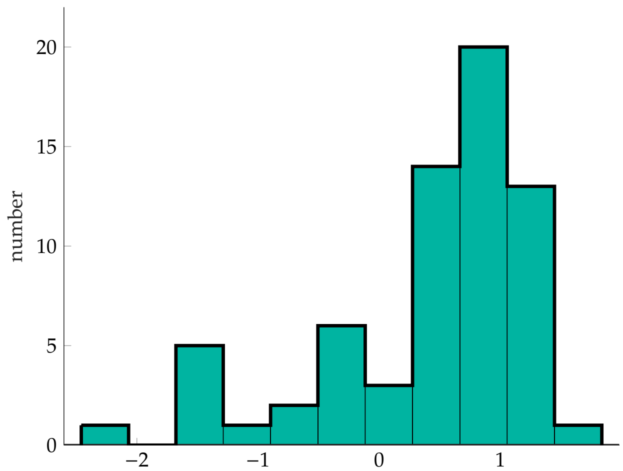
For this exercise, we focus on the IRS case where $\gamma = 0.875$. We consider two alternative calibrations: one where we set markups equal to their CES monopolistic values, and one where we set markups equal to their estimated values. We begin by discussing the case where all markups are set equal to their CES Dixit and Stiglitz (1977) values. Then we



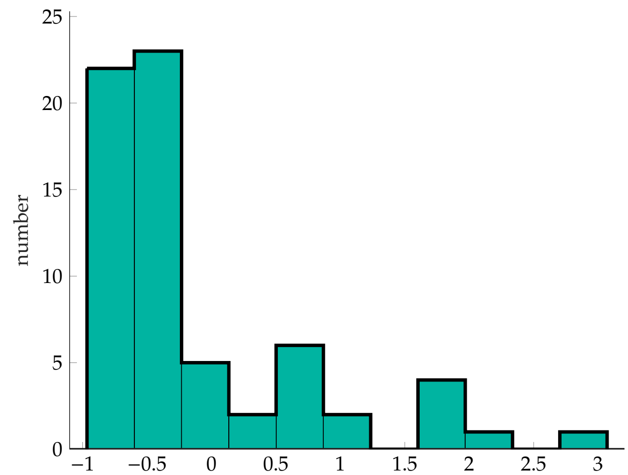
(a) Markup reduction for CES markups



(b) Entry subsidies for CES markups



(c) Markup reduction for estim. markups



(d) Entry subsidies for estim. markups

Figure 6: The elasticity of output with respect to reductions in markups or an entry subsidy to different sectors normalized by the cost of the intervention. The top row uses CES markups, whereas the bottom row uses estimated PF markups.

discuss the case where markups are equal to their estimated values in the data. In both cases, we abstract from endogenous changes in markups in response to the policy.[6]

The monopolistic-markups calibration is a useful starting point for understanding the results, since by setting markups to be the same in every sector, it helps isolate the role played by the input-output network on its own. In this case, markup reductions, plotted in Figure 6a, are always beneficial. Because we have imposed the same love-of-variety parameter in all sectors, the greatest bang-for-buck comes from reducing markups for those sectors with more complex supply chains, namely manufacturing industries like

[6]Here, we assume that the policy maker can directly change the wedges. As pointed out by Gupta (2020), in practice, a linear tax may not be able to achieve this since firm-level wedges may respond to the policy instrument.

motor vehicles, metals, and plastics. Intuitively, reducing markups in these sectors allows more entry into their supply chains. The smallest gains come from those industries with the simplest supply chains, mostly service industries like housing or legal services but also primary industries like oil extraction or forestry. For entry subsidies, plotted in Figure 6b, the biggest gains, on the other hand, come from subsidizing those industries which are upstream in complex supply chains, namely primary industries like forestry, oil, and mining, whereas subsidizing entry into relatively downstream industries, like nursing, hospitals, or social assistance, is actually harmful.

When we move to the estimated markups, plotted in Figures 6c and 6d, the shape of the input-output network is not the only determinant of the relative ranking of different industries, as now we must also consider whether each sector's markups are too high or too low on average relative to its external economies. Since, for simplicity, we have imposed the same love-of-variety effect in all sectors but we have estimated markups for each sector, we do not read too much into the exact relative ranking of the different industries.

However, these figures are still useful because they show that as we move farther away from the efficient frontier, which we do when we go from monopolistic markups to estimated markups, the potency of second-best policies increases dramatically. To see this, note that the elasticities in the top row are an order of magnitude smaller than the elasticities in the bottom row of Figure 6.

But the larger effect sizes are a mixed blessing. Once we are far away from the frontier, the scope for policy having unintended consequences also increases. Although there appear to be many free lunches available to policy makers, interventions can equally have large negative as well as positive effects. In other words, as implied by the theory of the second-best, interventions that seem sensible in isolation, like reducing markups, can reduce output once we are deep inside the frontier.

Appendix J Comparison to Simplified Models

Our analysis contends that careful modelling of the details of the production network and the entry technology is qualitatively important. To illustrate this quantitatively, in Table 3, we compare the results of the benchmark model to simplified versions of the model that employ some commonly used shortcuts: ignoring intermediate goods in production or entry (assuming no input-output); using a single-sector economy but allowing for intermediates (roundabout economy); ignoring firm-level heterogeneity within sectors (no firm heterogeneity). We discuss each of these strawmen in turn.

IRS	No Entry	Entry Uses Factors	Entry uses Goods/Factors
Benchmark	36%	50%	40%
No Input-Output	16%	20%	–
Roundabout	139%	182%	133%
Homogeneous Firms	4.6%	14%	10%
DRS			
Benchmark	26%	35%	32%
No Input-Output	13%	18%	–
Roundabout	91%	123%	108%
Homogeneous Firms	1.0%	7.8%	7.6%

Table 3: Efficiency losses from misallocation when different disaggregated aspects of the economy are trivialized. We use firm-level returns to scale $\varepsilon = 0.875$ under DRS, and $\gamma = 0.875$ under IRS. For IRS, this corresponds to an elasticity of substitution across firms within industries of 8.

The "No Input-Output" economy assumes away intermediates, and calibrates the size of each industry to be equal to its value-added share. Without entry, this economy undershoots the benchmark model for reasons discussed by Jones (2011) or Baqaee and Farhi (2019a). The undershooting becomes even more extreme once we allow for entry, underscoring even more strongly the need to model input-output linkages.

The "Roundabout" economy assumes that all firms in the economy belong to a single sector. The output of this sector is used both as the consumption good and as an intermediate input into production. This is a commonly used shortcut for incorporating intermediate inputs into a model. The one-sector roundabout economy overshoots the benchmark by a large amount. This is to be expected since the roundabout economy aggregates all firms in the economy into a single sector. This means cross-sectoral dispersions in markups (which are less costly than within-sectoral dispersions) are treated as if they are within-sectors. Intuitively, dispersed markups now distort input choices across producers by more, since firms in two different industries are treated as if they are highly substitutable.

Finally, the "Homogeneous Firms" economy assumes that all firms in a sector are identical, with the same productivity shifter and the same markup equal to the sectoral markup. The homogeneous sectors economy undershoots the benchmark by a large amount because even though it accounts for cross-sectoral distortions, it abstracts away from within-sector misallocation.

All in all, the sensitivity of these numbers underscores the quantitative importance of modelling and measuring the details as best we can.