Returns to Scale & Aggregate Productivity

Joel Kariel [1] Anthony Savagar [2]

[1]Competition and Markets Authority & University of Kent

[2]University of Kent

20 October 2023
CompNet, Brussels

Disclaimers

CMA

The work here does not represent the views of the CMA.

Data

This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

Funding

This research is funded under:

- ESRC project reference ES/V003364/1

Introduction

Motivation

- How can we understand rising returns to scale (RTS) and stagnating productivity?
- Growing evidence of rising returns to scale.
- Intuitively, recent technologies increase returns to scale:
  - intangible investment, IT and cloud infrastructure
- But as RTS rise in countries associated with these technologies, productivity is stagnating. This is puzzling...
- Typically higher RTS means higher productivity
  - True at the firm level
  - Unclear in aggregate due to competing channels e.g. selection, market power, allocation

What do we do? What do we find?

Empirical Contribution

- Estimate RTS in the UK economy.
- We find that firm-level RTS are increasing on average.

Theoretical Contribution

- How are productivity and RTS related?
- Firm dynamics model with imperfect competition and RTS.
  - Derive endogenous RTS and clarify sources of RTS.
  - Derive aggregate TFP and decompose
  - Show different sources of RTS have different implications
  - RTS in variable inputs seem better candidate than FC.
- Increases in RTS over the past couple of decades should've caused productivity to explode! **But**, if we allow markups to rise at the same time it erodes nearly all the productivity gains.

Returns to Scale

- Growing returns to scale:
  - Costs are less responsive to output.
  - Output is more responsive to inputs.
- **Cost side:** $\mathrm{RTS} = (d \ln \mathcal{C} / d \ln y)^{-1} = AC/MC$
- High RTS means costs unresponsive to output.
- Small firm has high RTS because fixed cost dominates, so change in output has little effect on costs.
- **Production side:** $\mathrm{RTS} = \nabla y(X) = \sum_{i=1}^{M} \frac{\partial \ln y}{\partial \ln x_i}$
- Output responds proportionally more for a small firm.

Returns to Scale Diagram

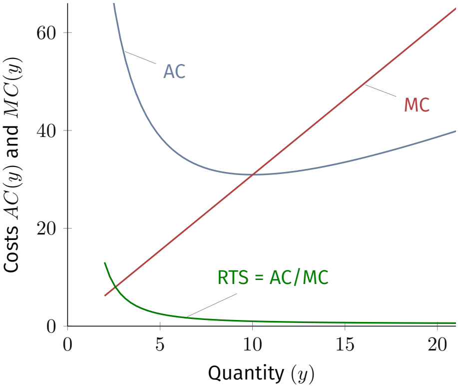


Figure 1 Fixed Cost with Increasing MC, U-Shaped AC Curve

Empirical Section

Data

- ARDx dataset is the UK's annual production survey.
- Runs from 1998 - 2014 and covers all sectors of the economy.
- 50,000 firms per year, 11m workers, 2/3 of GVA.
- All large firms ($>$250 employees) and a representative sample of smaller firms.
- We use data on: value added (revenue), labour (no. employees), materials and investment.
- We construct capital stock using the perpetual inventory method from firm-level investment data.

Returns to Scale Estimation

- We estimate Cobb-Douglas production functions:

$$\ln y_t(\imath) = \ln A_t(\imath) + \beta_1 \ln k_t(\imath) + \beta_2 \ln \ell_t(\imath) + \varepsilon_t(\imath).$$

- Endogeneity problem: cannot observe productivity $A_t(\imath)$ which affects optimal k_t, ℓ_t choices.
- Use Ackerberg, Caves, and Frazer (2015) and Gandhi, Navarro, and Rivers (2020) methodology.
- The sum of the coefficients is returns to scale (in variable inputs):

$$\beta_1 + \beta_2 = \nu.$$

- This measures the slope of the marginal cost curve.

Average Returns to Scale are Rising

- Pooling all firms over all years:
  - $\nu = 1.05$ and N = 527,813
- Pooling all firms over all years and studying sub-periods:

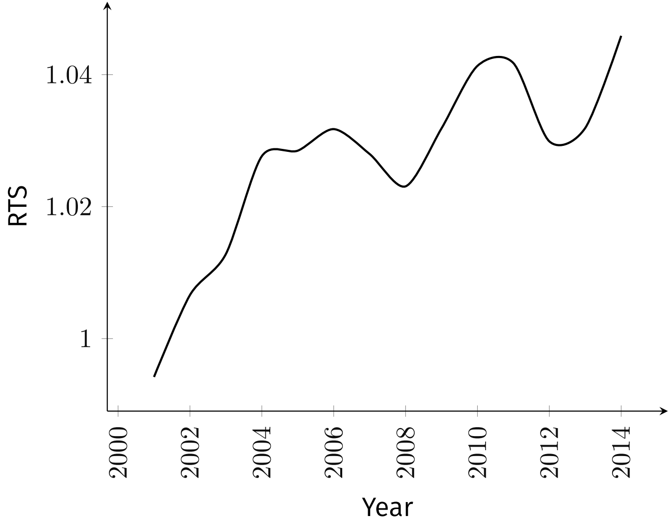
Table 1 Cobb-Douglas production function, 1998 - 2014, ACF

	1998 - 2001	2002 - 2005	2006 - 2009	2010 - 2014
ν	0.99	1.08	1.05	1.06
N	153,874	144,465	108,619	120,855

- At 2-digit industry vast majority experience increase.

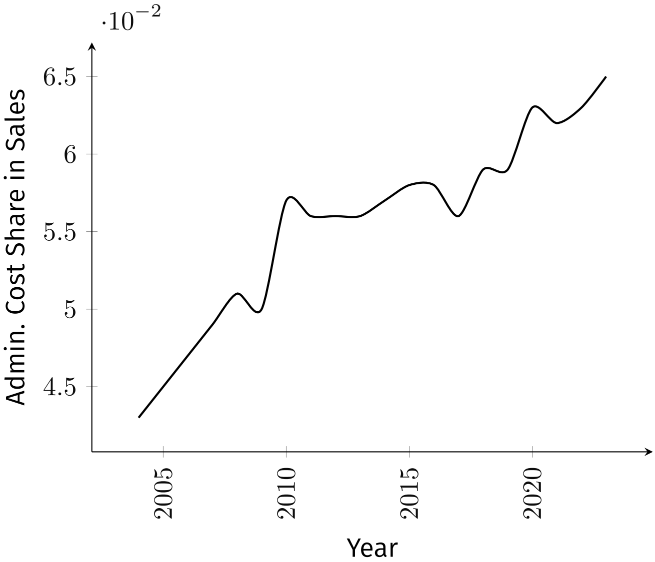
Rising Returns to Scale (rolling window GNR)

Figure 2 Returns to Scale in the UK, 2001 - 2014 using GNR



Fixed Cost Share is rising

Figure 3 Median Fixed Cost Share in Sales, BvD FAME



Model *(lite)*

Model Overview

- General equilibrium neoclassical growth with endogenous industry structure
- Monopolistic competition (fixed markups)
- Firm-level RTS
- Entry à la Hopenhayn (Pareto distributed firms)
- Productivity cut-off determining selection – how this moves is key.
- Household side standard
- Aggregate labour is fixed and exogenous.
- Firm side more involved...

Final Goods Producer

- Final goods producer solves

$$\Pi_t^F = \max_{y_t(\imath)} \quad Y_t - \int_0^{N_t} p_t(\imath) y_t(\imath) d\imath$$

$$\text{s.t.} \quad Y_t = N_t \left[\frac{1}{N_t} \int_0^{N_t} y_t(\imath)^{\frac{1}{\mu}} d\imath \right]^{\mu}$$

Markup

- The parameter $\mu \in (1, \infty)$ captures product substitutability.
- Optimality yields inverse-demand for firm:

$$p_t(\imath) = \left(\frac{Y_t}{N_t y_t(\imath)} \right)^{\frac{\mu - 1}{\mu}} .$$

- Hence there is downward-sloping demand.

Intermediate Goods Producer

Timeline

1. A firm pays cost κ to enter. Free entry holds.
2. Receives technology draw $A(\jmath)$ where $\jmath \in [0, 1]$ is uniform.
3. Given productivity draw, it decides whether to be active
   - Overhead cost ϕ causes some entrants to remain inactive.
4. All firms, both active and inactive, exit after one period.

Intermediate Goods Producer

- Firms have a fixed cost and sloping marginal costs

$$y_t(\jmath) = A(\jmath)\left[k_t(\jmath)^{\alpha}\ell_t(\jmath)^{1-\alpha}\right]^{\nu}$$
$$\ell_t(\jmath) = \ell_t^{tot}(\jmath) - \phi$$

- Therefore

$$y_t(\jmath) = A(\jmath)\left[k_t(\jmath)^{\alpha}[\ell_t^{tot}(\jmath) - \phi]^{1-\alpha}\right]^{\nu}$$

- Key parameters
  - ϕ fixed cost
  - ν RTS in variable inputs (MC slope) $\nu \uparrow \implies MC \downarrow$
- True RTS are a function of ϕ and ν
- Parameter ν captures *returns to scale in variable production*
  - $0 < \nu < 1$ decreasing returns in variable production (up-sloping MC)
  - $\nu = 1$ constant returns in variable production (flat MC)
  - $\nu > 1$ increasing returns in variable production (down-sloping MC)

Aggregate TFP

- $J_t \in (0,1)$ is productivity cut-off (i.e. the min prod. draw to cover your ϕ and make zero profits)
- E_t is measure of entrants $N_t = (1-J_t)E_t$
- Utilization u_t as the fraction of production labour in total labour:

$$u_t \equiv \frac{E_t \int_{J_t}^1 \ell(\jmath)d\jmath}{L_t} \qquad 1-u_t = \frac{N_t\phi}{L_t}.$$

- We can derive aggregate output

$$Y_t = TFP(J_t)K_t^{\alpha\nu}L_t^{1-\alpha\nu}$$

- TFP term captures aggregate productivity and can be decomposed into two terms $TFP_t = \Omega_t \times \hat{A}_t$:

$$TFP(J_t) \equiv \underbrace{\left(\frac{1-u_t}{\phi}\right)^{1-\nu} u_t^{(1-\alpha)\nu}}_{\text{allocative efficiency } \Omega_t} \underbrace{\left[\frac{1}{1-J_t}\int_{J_t}^1 A(\jmath)^{\frac{1}{\mu-\nu}}\, d\jmath\right]^{\mu-\nu}}_{\text{technical efficiency } \hat{A}_t}$$

- Alloc. eff. is all about fixed cost. Tech. eff. is selection.
- If $\phi = 0$ then $u_t = 1$ and $TFP_t = \hat{A}_t$

Model Analysis

Pareto Distribution

- Pareto distribution:

$$A(\jmath) = \frac{1}{(1-\jmath)^{1/\vartheta}}.$$

Pareto shape parameter

- $\vartheta > 1$ and as $\vartheta \to 1$ implies fatter right tail.
- An increase in J_t increases average productivity.
- The power mean, which appears in TFP, is:

$$\left[\frac{1}{1-J_t}\int_{J_t}^{1} A(\jmath)^{\frac{1}{\mu-\nu}}\, d\jmath\right]^{\mu-\nu} = \left(\frac{(\mu-\nu)\vartheta}{(\mu-\nu)\vartheta-1}\right)^{(\mu-\nu)} A(J_t)$$

- Technology term captures selection

Returns to Scale

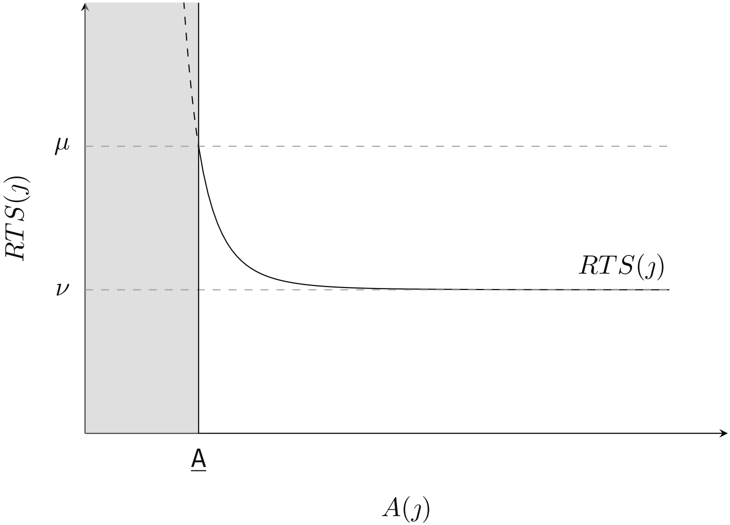
- Response of firm output to a change in *all* inputs:

$$\mathsf{RTS}_t(\jmath) \equiv \frac{\partial \ln y_t(\jmath)}{\partial \ln k_t(\jmath)} + \frac{\partial \ln y_t(\jmath)}{\partial \ln \ell_t^{tot}(\jmath)}$$
$$= \nu \left(1 + (1-\alpha)\frac{\phi}{\ell_t(\jmath)}\right)$$
$$= \nu + (\mu - \nu)\left[\frac{A(J_t)}{A(\jmath)}\right]^{\frac{1}{\mu-\nu}} .$$

- $\mathrm{RTS} \in (\nu, \mu)$ for high and low productivity draws.

Endogenous Returns to Scale

Figure 4 Firm-level Returns to Scale



Theoretical Results

- How do different sources of firm-level returns to scale (ν and ϕ) affect:
  1. Average returns to scale
  2. Average productivity

Average RTS Results

- Average firm under Pareto has

$$R\bar{T}S(J) = \nu + \frac{\vartheta(\mu-\nu)^2}{1+\vartheta(\mu-\nu)}$$

- Average returns to scale are increasing in span of control ν.

$$\frac{\partial R\bar{T}S(J)}{\partial \nu} = \frac{1}{(1+\vartheta(\mu-\nu))^2} > 0$$

- Average returns to scale are invariant to the fixed costs ϕ.

$$\frac{\partial R\bar{T}S(J)}{\partial \phi} = 0$$

- If ϕ increases:
  - All firms higher RTS.
  - But some high RTS firms become inactive.
  - Selection exactly offsets individual firm effect.

Aggregate Productivity

- Under Pareto model has a tractable steady state.
- *How do firm-level returns to scale affect aggregate productivity?*
  - Increase in ν
  - Increases in ϕ
- Aggregate productivity depends on allocative and technical efficiency

$$\ln TFP = \underbrace{\ln \Omega}_{alloc.} + \underbrace{\ln \hat{A}}_{tech.}$$

- Technical efficiency depends on selection.
- Allocative efficiency depends on the number of firms.

Change in ϕ

- Changes in fixed costs affect aggregate TFP through an allocation component and a technology component:

$$\frac{d\ln TFP}{d\ln\phi} = \frac{d\ln\Omega}{d\ln\phi} + \frac{d\ln\hat{A}}{d\ln\phi}$$

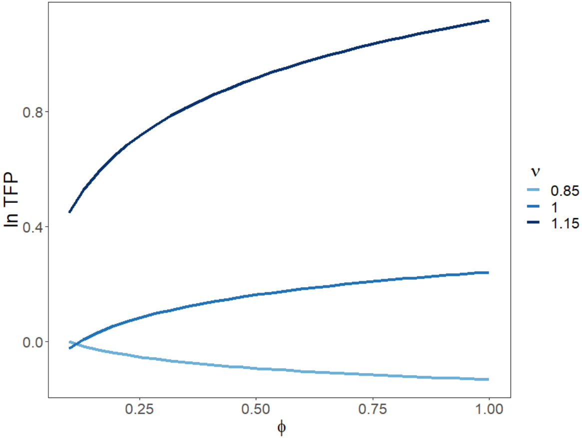
- Higher fixed costs increase selection which enhances productivity
- Allocation effect is ambiguous whether more or less firms is good for the division of resources depends on ν

$$\frac{d\ln TFP}{d\ln\phi} = -(1-\nu) + \frac{\nu(1-\alpha)}{\vartheta(1-\alpha\nu)-1}$$

- The effect is independent of μ

ϕ may raise productivity

Figure 5 Effect of ϕ on TFP for different ν



Change in ν

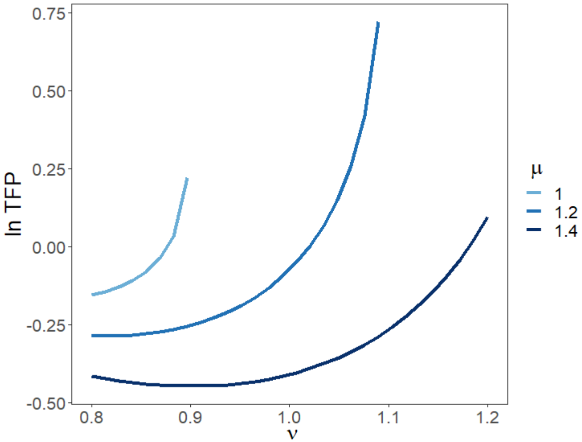
- TFP expression is nonlinear in ν

	Parameter	Value	Target
β	Discount rate	0.96	Real interest rate
δ	Depreciation rate	0.08	Office for National Statistics
ν	Variable RTS	0.99 - 1.05	ABS (authors' estimates)
μ	Markup	1.21 - 1.28	CMA (2022)
α	Capital share	0.25	ABS (authors' calculations)
ϑ	Pareto shape	10	Hopenhayn (2014)
κ	Entry cost	0.03	Match share inactive firms
ϕ	Overhead cost	0.2	Match share inactive firms

ν raises productivity

- Higher markups have a levels effect
- Weaken pass-through of RTS to TFP – mainly due to weaker selection decomposition

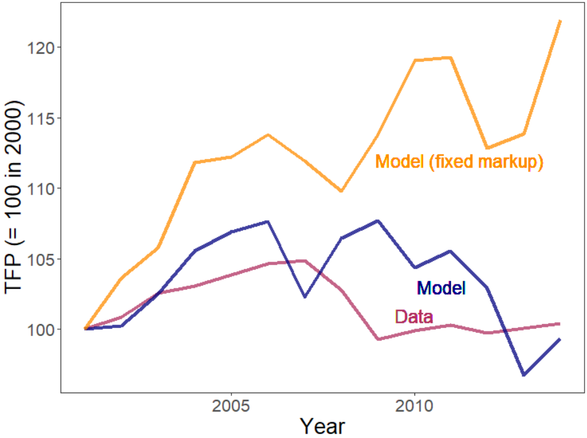
Figure 6 Effect of ν on TFP for different μ



Both ν and μ matter!

Calibrated Model

Figure 7 TFP Growth: Model vs Data



The TFP data series is from FRED.

Summing Up

- How can we understand rising returns to scale and stagnating productivity?
- Empirical results confirm returns to scale have increased whilst productivity has stagnated.
- Absent of markup increases productivity should've increased 20% over last two decades through RTS rises.
- However markup increases wipe out all these gains.
- RTS can increase through different sources – technologies that have reduced MC rather than increased FC seem more plausible.

References I

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer (2015). "Identification Properties of Recent Production Function Estimators". In: *Econometrica* 83.6, pp. 2411–2451.

Asturias, Jose, Sewon Hur, Timothy J. Kehoe, and Kim J. Ruhl (2022). "Firm Entry and Exit and Aggregate Growth". In: *American Economic Journal: Macroeconomics*.

Barseghyan, Levon and Riccardo DiCecio (Oct. 2011). "Entry costs, industry structure, and cross-country income and TFP differences". In: *Journal of Economic Theory* 146.5, pp. 1828–1851.

CMA (Apr. 2022). *The State of UK Competition*. Tech. rep.

Gandhi, Amit, Salvador Navarro, and David A. Rivers (2020). "On the Identification of Gross Output Production Functions". In: *Journal of Political Economy* 128.8, pp. 2973–3016.

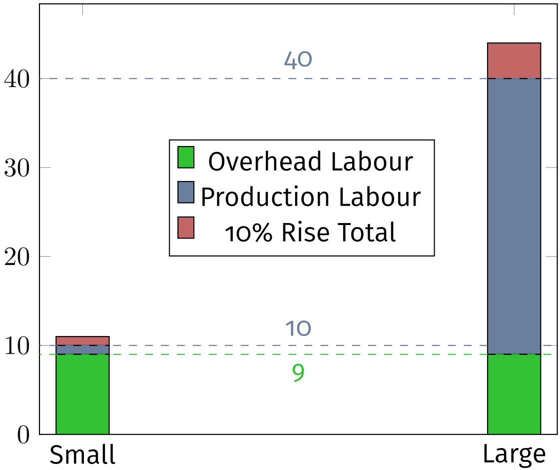
Hopenhayn, Hugo (2014). "Firms, misallocation, and aggregate productivity: A review". In: *Annual Review of Economics* 6.1, pp. 735–770.

References II

Luttmer, Erzo G. J. (Aug. 2007). “Selection, Growth, and the Size Distribution of Firms”. In: *The Quarterly Journal of Economics* 122.3, pp. 1103–1144.

Returns to Scale and Firm Size (Production Side)

Intuition – Why do small firms have greater RTS?



- 10% rise in **total labour** raises **production labour** by 100% for small firm, just 13% for large firm.
- Small firm has greater returns to scale because output more responsive to inputs.

Pareto Productivity Distribution I

Productivity $A(\jmath)$ is a random draw on the unit interval $\jmath \in [0,1]$ using inverse transform sampling. The Pareto CDF is given by

$$F(A;\vartheta) = 1 - \left(\frac{h}{A}\right)^{\vartheta}; \quad A \geq h > 0 \quad \text{and} \quad \vartheta > 0.$$

If $\mathcal{J} \sim Uniform(0,1]$, then for $\jmath \in \mathcal{J}$, we have

$$1 - \left(\frac{h}{A}\right)^{\vartheta} = \jmath$$

Therefore

$$A(\jmath) = h(1-\jmath)^{-\frac{1}{\vartheta}}.$$

We set the scale parameter – which is the minimum possible value of A – to $h = 1$. Calibrations of the shape parameter (tail index) are set to match the firm size distribution, for example $\vartheta = 1.15$ in Barseghyan and DiCecio (2011) and $\vartheta = 1.06$ in Luttmer (2007) and $\vartheta = 6.10$ in Asturias, Hur, Kehoe, and Ruhl (2022).

Pareto Productivity Distribution II

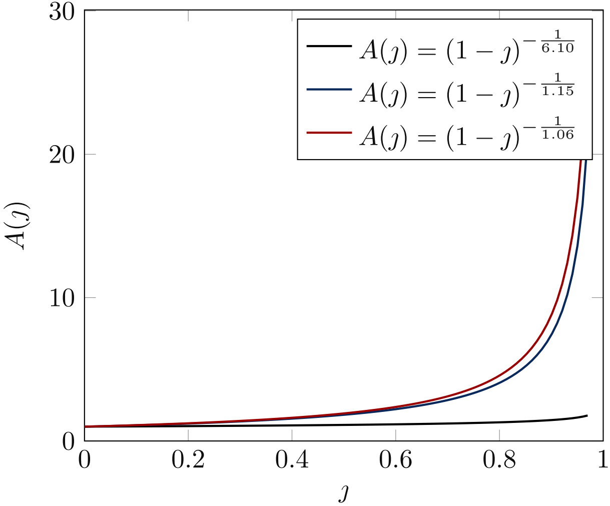
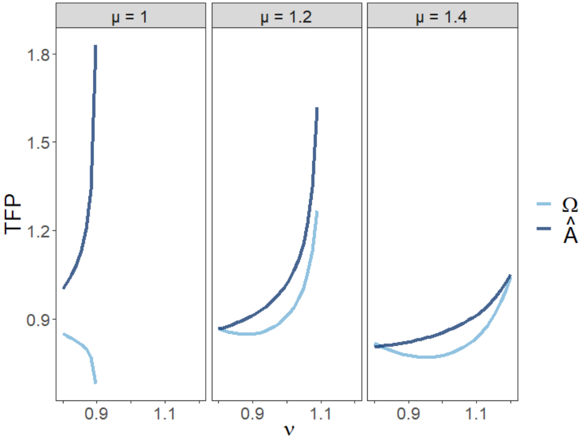


Figure 8 Productivity with Pareto Distribution, $h = 1, \vartheta = \{1.06, 1.15, 6.10\}$

Markups weaken the selection effect of higher RTS

- Higher markups weaken the selection channel which limits the productivity gains from higher RTS.
- Higher RTS increase productivity through greater selection but that channel is weakened when markups go up

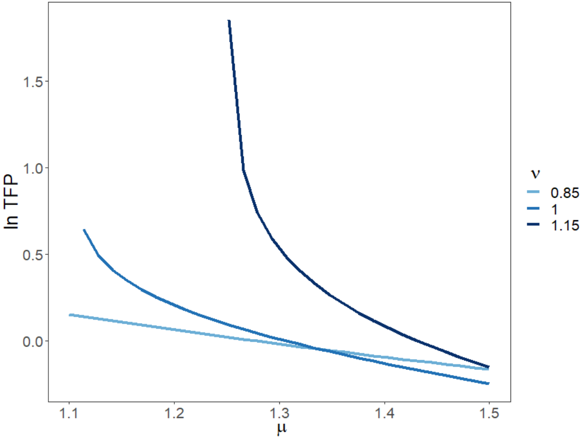
Figure 9 Effect of variable RTS on decomposed TFP for different markups



μ lowers productivity

- Higher markups reduce TFP (weaker selection)
- Rising markups are made worse by higher returns to scale

Figure 10 Effect of μ on TFP for different ν



Full Model

Household

- Household solves

$$
\begin{aligned}
\max_{\{C_t, K_{t+1}\}_{t=0}^{\infty}} \quad & \sum_{t=0}^{\infty} \beta^t U(C_t), \quad \beta \in (0,1), \\
\text{s.t.} \quad & C_t + I_t = r_t K_t + w_t L^{\mathrm{s}} + \Pi_t + T_t \\
& I_t = K_{t+1} - (1-\delta) K_t.
\end{aligned}
$$

- $L^{\mathrm{s}} = 1$ normalise labour supply.
- T_t are entry costs that government rebates to households.
- Π_t are total profits. Revenue less factor payments less entry costs.
- Optimality condition is

$$
\left(\frac{C_{t+1}}{C_t}\right)^{\sigma} = \beta(r_{t+1} + (1-\delta)).
$$

Factor Market Equilibrium

- Intermediate goods producer solves

$$\pi_t(\jmath) = \max_{k_t(\jmath), \ell_t(\jmath)} p_t(\jmath) y_t(\jmath) - r_t k_t(\jmath) - w_t(\ell_t(\jmath) + \phi)$$

- Subject to production function and inverse demand.
- Results in factor market equilibrium:

$$\begin{aligned} \frac{r_t}{p_t(\jmath)} &= \frac{\nu}{\mu} \alpha \frac{y_t(\jmath)}{k_t(\jmath)} \\ \frac{w_t}{p_t(\jmath)} &= \frac{\nu}{\mu}(1-\alpha) \frac{y_t(\jmath)}{\ell_t(\jmath)}. \end{aligned}$$

- Real factor prices equal to marginal revenue products of input.
- Firms charge markup $\mu \in (1, \infty)$ of price over marginal cost.

Free Entry

- All firms die after one period.
- A firm only produces if it makes positive profits, hence firm value is given by
$$v_t(\jmath) = \max\{\pi_t(\jmath), 0\}.$$
- Free entry condition implies that the expected value of a firm equals the entry cost
$$\int_0^1 v_t(\jmath) d\jmath = \kappa.$$

Firm size ratio

- From factor market equilibrium and inverse demand function.
- For two firms $\imath$ and $\jmath$ the ratio of firm size equals the scaled productivity ratio:

$$\frac{p_t(\jmath)y_t(\jmath)}{p_t(\imath)y_t(\imath)} = \frac{k_t(\jmath)}{k_t(\imath)} = \frac{\ell_t(\jmath)}{\ell_t(\imath)} = \left[\frac{A_t(\jmath)}{A_t(\imath)}\right]^{\frac{1}{\mu-\nu}} \quad \forall \imath, \jmath,$$

Zero-Profit Productivity Threshold

- Given factor market equilibrium, profits are

$$\pi(\jmath) = \left(1 - \frac{\nu}{\mu}\right) p_t(\jmath) y_t(\jmath) - w_t \phi$$

- At productivity draw $J_t \in (0, 1)$ firm makes zero profit

$$\left(1 - \frac{\nu}{\mu}\right) p_t(J_t) y_t(J_t) - w_t \phi = 0.$$

- Productivity draw $\jmath \in (0, J_t)$ firm inactive; $\jmath \in (J_t, 1)$ firm active
- $J_t \uparrow$ stronger selection . $J_t \downarrow$ weaker selection.
- Expected profits conditional on surviving are

$$\mathbb{E}[\pi_t] = (1 - J_t) w_t (J_t) \phi \left(\frac{1}{1 - J_t} \int_{J_t}^{1} \left[\frac{A(\jmath)}{A(J_t)} \right]^{\frac{1}{\mu - \nu}} d\jmath - 1 \right)$$

Aggregation I

- Operating firms are a fraction of entering firms

$$N_t = E_t \int_{J_t}^{1} d\jmath = E_t(1 - J_t)$$

- J_t is the fraction of inactive firms ("exit") or *selection*.
- $1 - J_t = N/E$ is fraction of active firms ("survival").
- Aggregate factor inputs

$$K_t = E_t \int_{J_t}^{1} k_t(\jmath)\, d\jmath$$

$$L_t = E_t \int_{J_t}^{1} \ell_t(\jmath) + \phi\, d\jmath$$

Aggregation II

- Utilization u_t as the fraction of production labour in total labour:

$$u_t \equiv \frac{E_t \int_{J_t}^{1} \ell(\jmath) d\jmath}{L_t} \qquad 1 - u_t = \frac{N_t \phi}{L_t}.$$

- We can derive aggregate output

$$Y_t = TFP(J_t) K_t^{\alpha\nu} L_t^{1-\alpha\nu}$$

- TFP term captures aggregate productivity and can be decomposed into two terms $TFP_t = \Omega_t \times \hat{A}_t$:

$$TFP(J_t) \equiv \underbrace{\left(\frac{1-u_t}{\phi}\right)^{1-\nu} u_t^{(1-\alpha)\nu}}_{\text{allocative efficiency } \Omega_t} \underbrace{\left[\frac{1}{1-J_t}\int_{J_t}^{1} A(\jmath)^{\frac{1}{\mu-\nu}}\ d\jmath\right]^{\mu-\nu}}_{\text{technical efficiency } \hat{A}_t}$$

- If $\phi = 0$ then $u_t = 1$ and $TFP_t = \hat{A}_t$

Closing the model

- The resource constraint

$$Y_t = C_t + I_t$$

- Entry fees are rebated to households by the government

$$T_t = E_t \kappa$$

- Profits and labour market clearing

$$\Pi_t = \Pi_t^F$$
$$L_t = L_t^s$$