Returns to Scale & Aggregate Productivity

Joel Kariel [1] Anthony Savagar [2]

[1]Competition and Markets Authority & University of Kent

[2]University of Kent

Compass Lexecon Economics Conference 2023

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

Returns to Scale

- Growing RTS:
  - Costs less responsive to output.
  - Output more responsive to inputs.
- **Cost side:** $\text{RTS} = (d \ln \mathcal{C} / d \ln y)^{-1} = AC/MC$
- **Production side:** $\text{RTS} = \nabla y(X) = \sum_{i=1}^{M} \frac{\partial \ln y}{\partial \ln x_i}$

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

Returns to Scale

- Growing RTS:
  - Costs less responsive to output.
  - Output more responsive to inputs.
- **Cost side:** $\mathrm{RTS} = (d \ln \mathcal{C} / d \ln y)^{-1} = AC/MC$
- **Production side:** $\mathrm{RTS} = \nabla y(X) = \sum_{i=1}^{M} \frac{\partial \ln y}{\partial \ln x_i}$

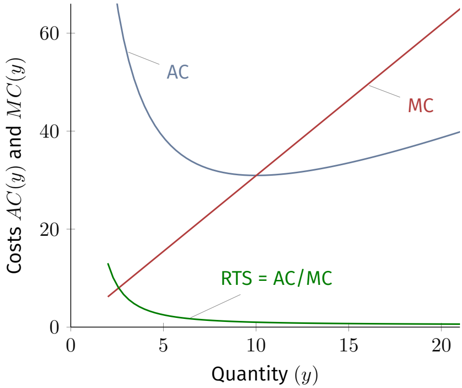


Figure 1 Fixed Cost with Increasing MC, U-Shaped AC Curve

Motivation

- Growing evidence of rising returns to scale (RTS).
- Intuitively, recent technologies ought to boost scale:
  - intangible investment, IT and cloud infrastructure
- But as RTS rises in the UK, productivity is stagnating. Why?
- Typically higher RTS means higher productivity.
  - In aggregate, depends on: **sources of RTS** and **firm selection**.

Intuition

- For profit-maximising firms:

$$\nu = \mu(1 - s_\pi)$$

- Interaction between RTS, market power, and firm survival.
- Technologies that enhance RTS may also raise market power.
- Higher markups chip away at the boost to aggregate productivity by allowing low efficiency firms to survive.

Intuition

- For profit-maximising firms:

$$\nu = \mu(1 - s_\pi)$$

- Interaction between RTS, market power, and firm survival.
- Technologies that enhance RTS may also raise market power.
- Higher markups chip away at the boost to aggregate productivity by allowing low efficiency firms to survive.

What do we do? What do we find?

Theoretical Contribution

- How are productivity and RTS related?
- Source of RTS matters! RTS depend on MC ν, FC ϕ, and (indirectly) the markup μ.
- Firm dynamics model with imperfect competition and endogenous RTS.

Empirical Contribution

- Are the model predictions consistent with the data?
- Estimate RTS and productivity in the UK economy using recent techniques in production function estimation.
- Rising RTS should raise aggregate productivity; higher market power can overturn this result.

Empirical Findings

What Do We Estimate?

- We estimate 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 coefficients are

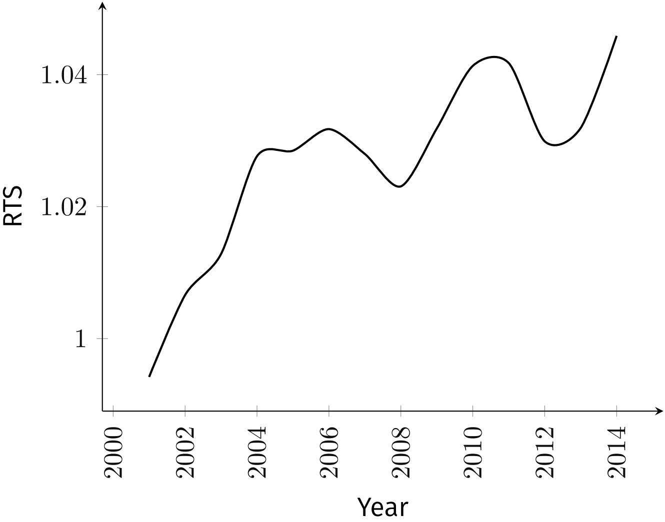
$$\beta_1 = \frac{\partial \ln y_t(\imath)}{\partial \ln k_t(\imath)} = \nu\alpha \qquad \beta_2 = \frac{\partial \ln y_t(\imath)}{\partial \ln \ell_t(\imath)} = \nu(1-\alpha).$$

- Sum of coefficients is:

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

Rising RTS

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



Higher Scale Across Sectors

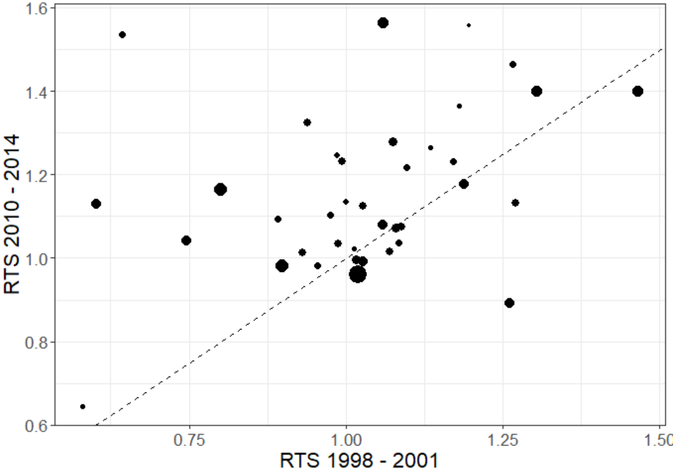


Figure 3 Sectoral RTS Estimates 1998-2001 vs 2010-2014, ACF

Model

Model Outline

- Imperfect competition with fixed markup $\mu \rightarrow$ downwards-sloping demand.
- Firm pays cost κ to enter and receives technology draw $A(\jmath)$.
- Firm decides to produce, incurring labour overhead ϕ.
- Free-entry and zero-profit condition control which firms survive. Cut-off firm has productivity denoted $\underline{\text{A}}$.
- Pareto distribution on $A(\jmath)$ with shape ϑ.
- Aggregate productivity:

$$\ln TFP = \ln \underbrace{\Omega}_{\text{Allocative}} + \ln \underbrace{\hat{A}}_{\text{Technical}}$$

Allocative Efficiency Technical Efficiency

Firms

- Firm production function given by

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

- Production labour is total labour less fixed overhead:

$$\ell_t(\jmath) = \ell_t^{\text{tot}}(\jmath) - \phi.$$

- True RTS are a function of ν, ϕ and μ.

Returns to Scale

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

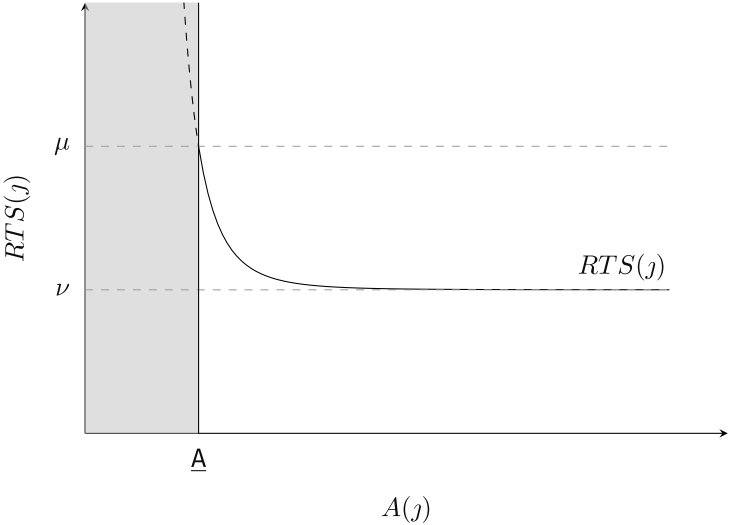
$$\mathsf{RTS}_t(\jmath) \equiv \frac{\partial \ln y_t(\jmath)}{\partial \ln k_t(\jmath)} + \frac{\partial \ln y_t(\jmath)}{\partial \ln \ell_t^{tot}(\jmath)}$$
$$= \nu \left(1 + (1-\alpha)\frac{\phi}{\ell_t(\jmath)}\right)$$
$$= \nu + (\mu - \nu)\left[\frac{\underline{\mathsf{A}}}{A(\jmath)}\right]^{\frac{1}{\mu-\nu}}.$$

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

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

Endogenous Returns to Scale

Figure 4 Firm-level Returns to Scale



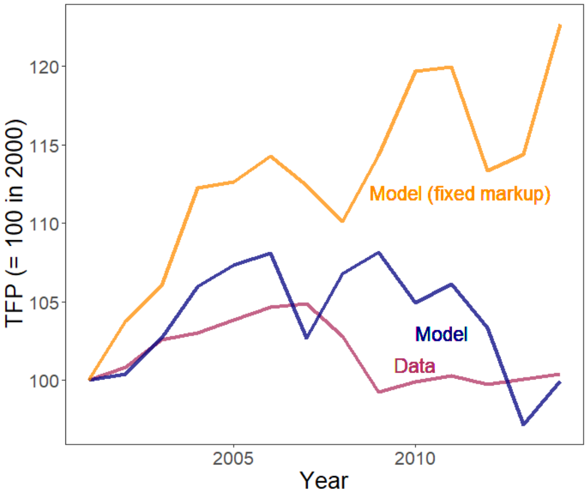
How do ν, ϕ, and μ affect RTS & TFP?

- Higher ν $\rightarrow$ raises RTS & aggregate productivity. Graph
- Higher ϕ $\rightarrow$ no effect on RTS & ambiguous effect on aggregate productivity. Graph
- Higher μ $\rightarrow$ reduces RTS & aggregate productivity. Graph

Both ν and μ matter!

Calibrated Model

Figure 5 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?
- RTS depends on marginal costs and fixed costs.
- Aggregate productivity rises with scale, but falls with markups.
- Empirical results confirm returns to scale have increased whilst productivity growth has slowed.
- **Take-away:** Technologies that have lowered marginal costs – not raised fixed costs – can enhance RTS + raise markups, and this combination yields stagnating productivity growth.

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.

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda (Dec. 2020). "Changing Business Dynamism and Productivity: Shocks versus Responsiveness". In: *American Economic Review* 110.12, pp. 3952–90.

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.

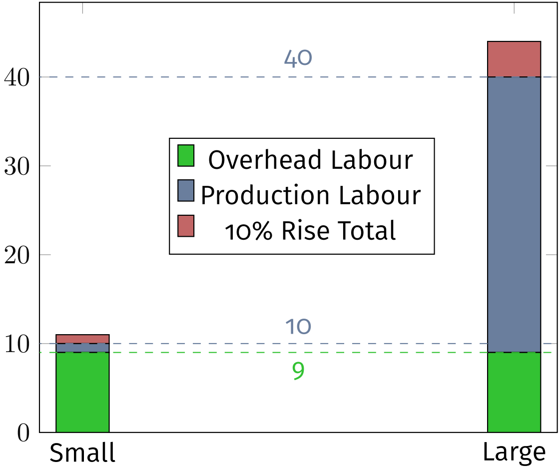
References II

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.

Melitz, Marc J (Nov. 2003). “The impact of trade on intra-industry reallocations and aggregate industry productivity”. In: *Econometrica* 71.6, pp. 1695–1725.

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.

Household I

- Household solves

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

- $L^{\mathrm{s}} = 1$ 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

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

Final Goods Producer I

- Final goods producer solves

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

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

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 I

Timeline

1. A firm pays cost κ to enter.
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.

Factor Market Equilibrium I

- 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 I

- 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 I

- 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 I

- 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

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

Average Productivity I

- Average technology of active firms:

$$\bar{A}(J) = \frac{1}{1 - J_t} \int_{J_t}^{1} A(\jmath) d\jmath$$

- The following power mean appears later:

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

- Generalised version of Melitz (2003) and many others.

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:

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

- We can derive aggregate output

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

- TFP captures aggregate productivity

$$TFP(J_t) \equiv \left(\frac{1-u_t}{\phi}\right)^{1-\nu} u_t^{(1-\alpha)\nu} \left[\frac{1}{1-J_t} \int_{J_t}^{1} A(\jmath)^{\frac{1}{\mu-\nu}} \; d\jmath\right]^{\mu-\nu}$$

Closing the model I

- 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

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

Allocative Efficiency

- Allocative efficiency is the allocation of resources across firms. It depends on the number of firms N and the share of labour used to produce output (rather than overheads) u.

$$\Omega = \left(\frac{N}{\phi}\right)^{(1-\nu)} u^{(1-\alpha)\nu} \qquad u = \left(1 + \frac{\vartheta(\mu - \nu) - 1}{\nu\vartheta(1-\alpha)}\right)^{-1}$$

- Number of firms $N \downarrow$ in ν, $\downarrow$ in ϕ, $\uparrow$ in μ.
- Share of production labour $u \uparrow$ in ν, unaffacted by ϕ, $\downarrow$ in μ.

Return

Technical Efficiency

- Technical efficiency is a power-mean of productivity for surviving incumbents.

$$\hat{A}(J) \equiv \left[\frac{1}{1-J} \int_J^1 A(\jmath)^{\frac{1}{\mu-\nu}} \, d\jmath \right]^{\mu-\nu} .$$

  where J is the productivity cut-off on the interval $[0,1]$.
- Under Pareto, it takes the form:

$$\hat{A} = \left(\frac{\vartheta(\mu-\nu)}{\vartheta(\mu-\nu)-1} \right)^{\mu-\nu} \underline{\mathsf{A}} = \Gamma \underline{\mathsf{A}}.$$

  where $\underline{\mathsf{A}}$ is the cut-off productivity.
- $\Gamma \uparrow$ in ν, unaffeced by ϕ, $\downarrow$ in μ.
- Productivity cut-off $\underline{\mathsf{A}} \uparrow$ in ν, $\uparrow$ in ϕ, $\downarrow$ in μ.

Return

Average RTS I

- Average firm under Pareto has

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

- As $\mu - \nu \to 0$ then $R\bar{T}S \to \nu = \mu$.
- Average returns to scale are increasing in span of control ν, decreasing in markup μ.

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

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

- Average returns to scale are invariant to the fixed costs ϕ.
  - All firms higher RTS.
  - But some high RTS firms become inactive.
  - Selection exactly offsets individual firm effect.

Productivity is Determined by Selection I

- Average productivity changes through the selection channel:

$$\frac{d\ln\bar{A}}{d\ln x} = \frac{\partial \ln\bar{A}}{\partial \ln(1-J)} \times \frac{d\ln(1-J)}{d\ln x}, \quad \text{where } x = \{\nu, \phi\}$$

- Average productivity is always decreasing in survival:

$$\frac{\partial \ln\bar{A}}{\partial \ln(1-J)} = -\frac{1}{\vartheta}$$

- More survival means less selection, so lower average productivity.

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

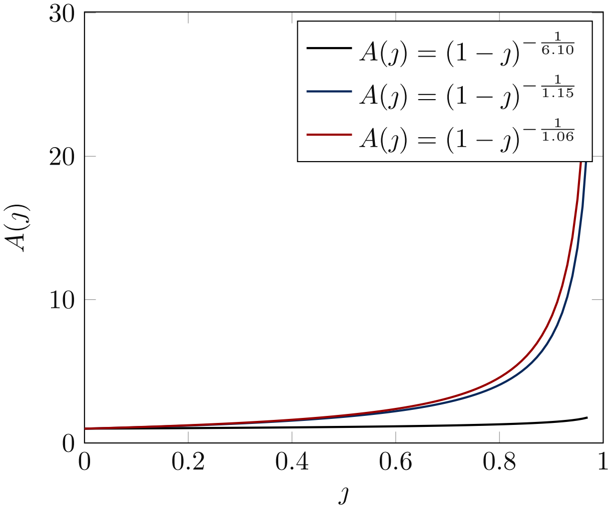


Figure 6 Productivity with Pareto Distribution, $h = 1, \vartheta = \{1.06, 1.15, 6.10\}$

Pareto Distribution

- Pareto distribution:

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

Pareto shape parameter

- $\vartheta > 1$ and as $\vartheta \to 1$ implies fatter right tail.
- The arithmetic average of technology is:

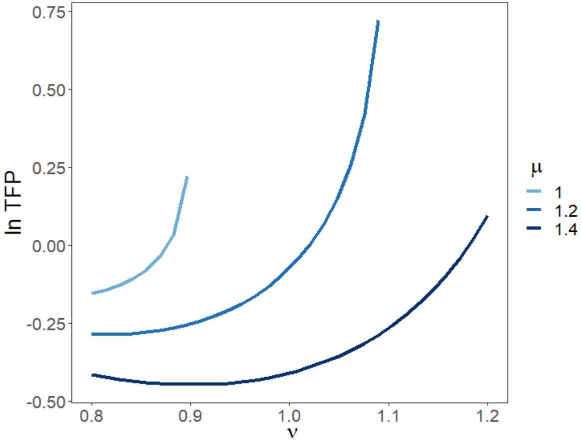
$$\bar{A}(J_t) = \frac{\vartheta}{\vartheta - 1} A(J_t)$$

- Average productivity is linearly related to threshold productivity.
- An increase in J_t increases average productivity.
- The power mean, which appears in TFP, is a function of $\bar{A}(J_t)$:

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

ν raises productivity

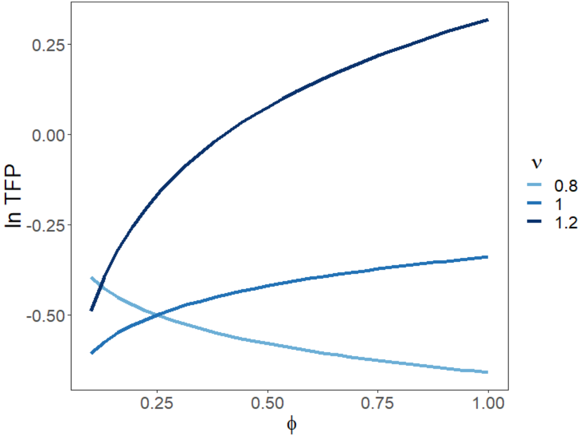
Figure 7 Effect of ν on TFP for different μ



Return

ϕ may raise productivity

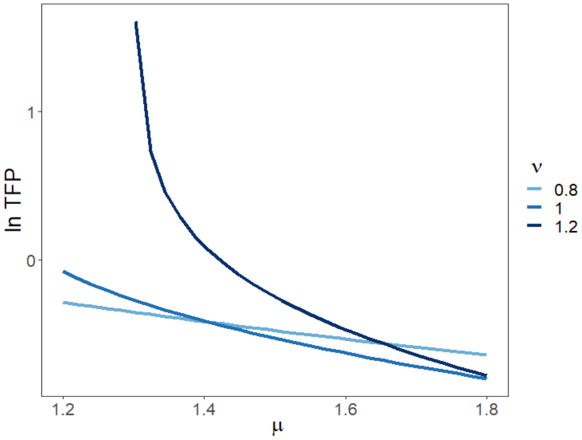
Figure 8 Effect of ϕ on TFP for different ν



Return

μ lowers productivity

Figure 9 Effect of μ on TFP for different ν



Return

Log-Normal Productivity Distribution I

We also consider the model with a log-normal productivity distribution. The log-normal CDF is given by:

$$F(A;\mu,\sigma) = \Phi\left(\frac{\ln A - \mu}{\sigma}\right); \quad A \geq 0 \quad \text{and} \quad \sigma > 0.$$

where Φ is the CDF of the standard normal distribution.
The PDF takes the equation:

$$f(A;\mu,\sigma) = \frac{1}{A\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln A - \mu)^2}{2\sigma^2}\right)$$

We cannot obtain closed-form solutions for expected profits as is possible with a Pareto distribution. Instead we use the following result to obtain the expectation of profits, conditional on the (scaled) productivity draw being above the cut-off a^*.

$$E[a \mid a \geqslant a^*] = e^{\mu + \frac{\sigma^2}{2}} \cdot \frac{\Phi\left[\frac{\mu+\sigma^2-\ln(a^*)}{\sigma}\right]}{1 - \Phi\left[\frac{\ln(a^*)-\mu}{\sigma}\right]}$$

Log-Normal Productivity Distribution II

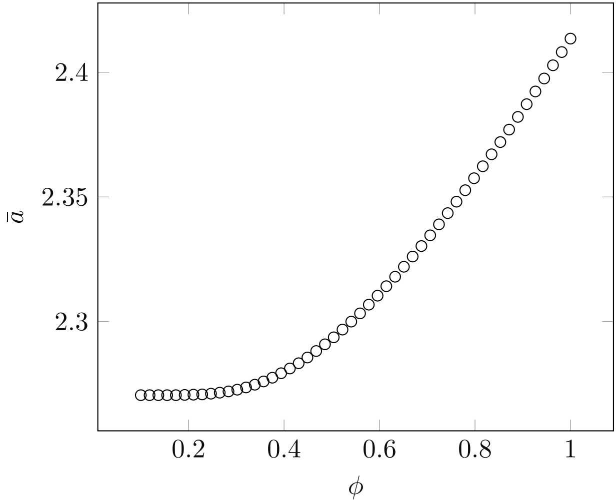
With this expression, the free-entry condition is written:

$$\kappa = w\phi \left(\frac{E[a \mid a \geqslant a^*]}{a^*} - 1 \right)$$

However, note that the wage w is an equilibrium object which depends on the productivity cut-off a^*. We choose a^* to solve the free-entry condition, recomputing the wage which depends on a^* at each iteration. With a solution to the productivity cut-off, we obtain average productivity $E[a \mid a \geqslant a^*]$ and average returns to scale $\nu + (\mu - \nu)\frac{a^*}{\bar{a}}$. In addition, we compute firm-level returns to scale $\nu + (\mu - \nu)\frac{a^*}{a(\jmath)}$ for any (scaled) productivity draw $a(\jmath)$.
Figures 10 and 11 show how changing ϕ affects average productivity and average returns to scale when productivity is log-normally distributed. Average productivity is rising in ϕ, as a higher fixed cost increases the productivity cut-off. In addition, average returns to scale rises in ϕ.

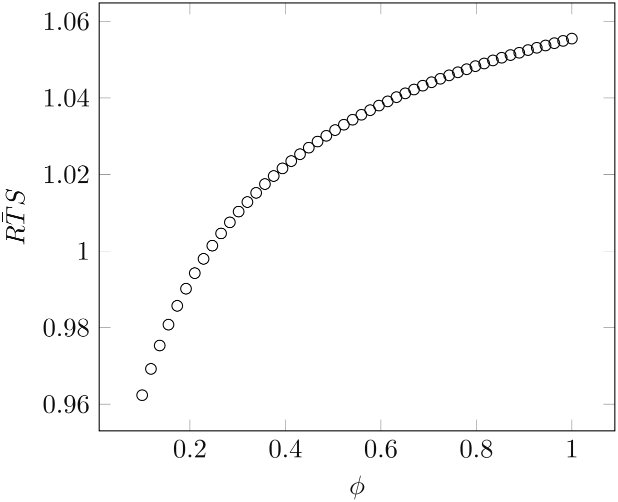
Log-Normal Productivity Distribution III

Figure 10 Average Productivity $\bar{a}(J)$ and Fixed Cost ϕ



Log-Normal Productivity Distribution IV

Figure 11 Average returns to scale $RTS(\bar{j})$ and Fixed Cost ϕ



TFPR as a proxy for technology

- Estimated TFPR, where hat notation is an estimated value, is:

$$\ln T\hat{FPR}_t(\jmath) = \ln p_t(\jmath) y_t(\jmath) - \hat{\beta}_1 \ln k_t(\jmath) - \hat{\beta}_2 \ln \ell_t(\jmath).$$

- In the model this residual is a composite of an aggregate demand shock and firm-level technology:

$$\ln TFPR_t(\jmath) = \frac{\mu - 1}{\mu} \ln(Y_t/N_t) + \frac{1}{\mu} \ln A_t(\jmath).$$

- Decker, Haltiwanger, Jarmin, and Miranda (2020, p. 3, 961)

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, labour (no. employees), materials and investment.
- We construct capital stock using the perpetual inventory method from firm-level investment data.