Endogenous Returns to Scale

Alexandr Kopytov
University of Rochester

Mathieu Taschereau-Dumouchel
Cornell University

Zebang Xu
Cornell University

Summer 2025

Intro

- Scalability of firms' technologies is important for their size
  - Modern superstar firms operate at massive scales
- Firm size is typically driven by its productivity
- Firms also differ in their technologies, in particular, returns to scale
- **Key premise of our paper**: firms have some control of their RTS
  - Run a local business vs invest in scalability to become international

What drives firms' RTS decisions? What are the aggregate implications?

Preview of the paper

- Multisector economy with within-sector heterogeneity
  - Firms have heterogeneous productivities and can adjust their RTS accordingly
  - High RTS allows firms to produce more but may come at a cost
  - Model allows for tractable aggregation
- Productive firms are large and have high RTS (in line with empirical evidence)
  - Endogenous RTS amplify size-productivity link
- Endogenous RTS amplify macro responses to changes in the environment
- Endogenous RTS dispersion increases GDP
  - Productivity dispersion makes endogenous RTS particularly valuable for GDP
- Calibration using Spanish data (in progress...)
  - We match within-sector dispersion in RTS and profits

Literature

- **Production function and RTS estimation**
  - Hall (1990), Burnside (1996), Basu and Fernald (1997), De Loecker et al. (2020), Gao and Kehrig (2020), Demirer (2020), Kariel and Savagar (2022), Ruzic and Ho (2023), Chiavari (2024), Chan et al. (2024)
- **Endogenous production scale**
  - Argente et al. (2021), Smirnyagin (2022), Lashkari et al. (2024), Chen et al. (2023), Hsieh and Rossi-Hansberg (2023), De Ridder (2024)
- **Technique choice in production networks**
  - Oberfield (2018), Acemoglu and Azar (2020), Kopytov et al. (2024)

I. Model

Environment

- Frictionless static model with competitive firms and representative household
- N sectors, with a continuum of firms in each sector
  - Firm l in sector i has DRS Cobb-Douglas production function

$$F_i\left(L_{il}, X_{il}, \eta_{il}\right) = e^{\varepsilon_{il} - a_i(\eta_{il})} \zeta_{il} \left(L_{il}^{1-\sum_{j=1}^{N} \alpha_{ij}} \prod_{j=1}^{N} X_{ij,l}^{\alpha_{ij}} \right)^{\eta_{il}}$$

  - $\varepsilon_{il} \sim$ iid $\mathcal{N}\left(\mu_i, \sigma_i^2\right)$ is productivity shock
  - Convex $a_i(\eta_{il})$ captures cost of operating high RTS technologies

▸ ζ

- Representative household owns the firms and supplies labor

$$\max \prod_{i=1}^{N} \left(\frac{C_i}{\beta_i}\right)^{\beta_i} \quad \text{s.t.} \quad \sum_{i=1}^{N} P_i C_i \leq W\bar{L}$$

  - Normalize price index: $\bar{P} = \prod_{i=1}^{N} P_i^{\beta_i} = 1$
  - Profits are dissipated through entry costs (more detail below)

Firm problem: Timing

1. Before ε is realized: Firms choose whether to enter

$$\text{[Free-entry condition]: } \mathbb{E}_i\left[\Pi_{il}\left(\varepsilon_{il}, P, W\right)\right] = \kappa_i W$$

2. After ε is realized: Firms choose quantities and returns to scale

$$\Pi_{il} = \max_{\eta_{il}, L_{il}, X_{il}} P_i F_i\left(L_{il}, X_{il}, \eta_{il}\right) - W L_{il} - \sum_{j=1}^{N} P_j X_{ij,l}$$

Lemma

Firm's marginal cost of production is

$$\lambda_{il} = \frac{1}{\exp(\varepsilon_{il} - a_i\left(\eta_{il}\right))} \Pi_{il}^{1-\eta_{il}} \left(W^{1-\sum_{j=1}^{N} \alpha_{ij}} \prod_{j=1}^{N} P_j^{\alpha_{ij}} \right)^{\eta_{il}},$$

where profit Π_{il} is the price of fixed entrepreneurial input.

Firm problem: Choice of RTS

- Returns to scale η_{il} is chosen to minimize marginal cost

$$\log\left(W^{1-\sum_{j=1}^{N}\alpha_{ij}}\prod_{j=1}^{N}P_j^{\alpha_{ij}}\right)-\log\Pi_{il}-\frac{da_i}{d\eta_{il}}=0$$

  - Increasing η_{il} shifts input mix from entrepreneurial factor to variable inputs bundle
  - It may also lead to a change in TFP directly

- Less expensive variable inputs $\Rightarrow$ higher η_{il}

- Any change pushing firm to be bigger (e.g., $\varepsilon_{il}\uparrow$ or $P_i\uparrow$) puts pressure on entrepreneurial factor which is in fixed supply $\Rightarrow$ firm relies less on it, i.e. η_{il} is higher

▸ Π_{il}

Lemma

Returns to scale η_{il} increases with productivity ε_{il} and price P_i but decreases with the price of variable input bundle $W^{1-\sum_{j=1}^{N}\alpha_{ij}}\prod_{j=1}^{N}P_j^{\alpha_{ij}}$.

Productivity, size, and RTS

Corollary

More productive firms (higher ε_{il}) earn higher profits, produce more, and pick higher RTS.

- In practice, ε_{il} is not observable
- Tornqvist index $\mathcal{T}_{i,kl}$ is commonly used to compare productivities of firms with different production functions (e.g., Penn World Table) Definition
  - $\mathcal{T}_{i,kl} > 1$ means that sector-i firm k is more productive than sector-i firm l

Lemma

Consider firms k and l in sector i with $\varepsilon_{ik} = \varepsilon_{il} + \Delta$. Then $\frac{d\mathcal{T}_{i,kl}}{d\Delta}\Big|_{\Delta=0} > 0$.

Equilibrium definition

Equilibrium definition

An *equilibrium* is a set of prices (P^*, W^*), a choice of returns to scale $\{\eta_{il}^*\}$, a tuple of firm-level quantities $\{C_{il}^*, L_{il}^*, X_{il}^*, Q_{il}^*\}$, and masses of firms in each sector $\{M_i^*\}$ such that

1. (Firm optimality) For each i and l, η_{il}^*, L_{il}^*, and X_{il}^* solve firm problem
2. (Consumer optimality) Consumption vector C^* solves household problem
3. (Free entry) For each i, $\int \Pi_i(\varepsilon_{il}, P^*, W^*)\, d\Phi_i(\varepsilon_{il}) = \kappa_i W^*$
4. Market clearing:

$$C_i^* + \sum_{j=1}^{N} X_{ji}^* = Q_i^* = \int_0^{M_i^*} F_i(L_{il}^*, X_{il}^*, \eta_{il}^*)\, dl \quad \text{and} \quad \sum_{i=1}^{N} L_i^* + \sum_{i=1}^{N} M_i^* \kappa_i = \bar{L}$$

II. Aggregation

Sectoral aggregation: Preliminaries

Assumption for tractable aggregation

The cost function takes the form $a_i(\eta_{il}) = \frac{\gamma_i}{1-\eta_{il}}$, where $\gamma_i > \sigma_i^2/2$.

- Define effective returns to scale in sector i as $\hat{\eta}_i = \left(\int \eta_{il} Q_{il} dl\right) / \left(\int Q_{il} dl\right)$
- Define effective productivity dispersion in sector i as $\varphi_i = \sigma_i^2/2\gamma_i \in [0,1)$

Lemma

Returns to scale of firm l in sector i can be expressed as

$$\frac{1}{1-\eta_{il}} = \frac{1-\varphi_i}{1-\hat{\eta}_i} + \frac{\varepsilon_{il}-\mu_i}{2\gamma_i}.$$

- If $\varphi_i > 0$, then $\hat{\eta}_i > \eta_i(\mu_i)$ because large firms have higher RTS

Sectoral aggregation

Proposition

Sectoral marginal cost of production is

$$\lambda_i = \frac{1}{\exp(z_i(\hat{\eta}_i))} (W\kappa_i)^{1-\hat{\eta}_i} \left(W^{1-\sum_{j=1}^{N}\alpha_{ij}} \prod_{j=1}^{N} P_j^{\alpha_{ij}} \right)^{\hat{\eta}_i},$$

where

$$z_i(\hat{\eta}_i) = \underbrace{\mu_i - (1-\varphi_i)\, a_i(\hat{\eta}_i)}_{\text{Productivity of the average firm}} - \underbrace{(1-\hat{\eta}_i)\log\sqrt{1-\varphi_i}}_{\text{Productivity increase due to dispersion}}.$$

- Cobb-Douglas sectoral marginal cost with three inputs: Labor, intermediate inputs, entry costs
- Sectoral productivity = productivity of the average firm + dispersion adjustment
  - High-ε_{il} firms pick high RTS and produce more at low average cost
  - If all firms pick $\eta_{il} = \hat{\eta}_i$, λ_i takes the same form but without dispersion adjustment

Prices and GDP

Proposition

1. In equilibrium, marginal costs equal prices:

$$\log\left(P/W\right) = -\mathcal{L}\left(\hat{\eta}\right)\left(z\left(\hat{\eta}\right) - \left(I - \operatorname{diag}\left(\hat{\eta}\right)\right)\log\kappa\right),$$

   where $\mathcal{L}\left(\hat{\eta}\right) = \left(I - \operatorname{diag}\left(\hat{\eta}\right)\alpha\right)^{-1}$ is the Leontieff inverse matrix.

2. Equilibrium log GDP is

$$y = \log(W\bar{L}) = \left[\omega\left(\hat{\eta}\right)\right]^{\top}\left(z\left(\hat{\eta}\right) - \left(I - \operatorname{diag}\left(\hat{\eta}\right)\right)\log\kappa\right) + \log\bar{L},$$

   where $\omega_i = \frac{P_i Q_i}{\bar{P}Y} = \beta^{\top}\mathcal{L}\left(\hat{\eta}\right)\mathbf{1}_i$ is Domar weight of sector i.

- RTS affects GDP through
  - Importance of different sectors, $\omega\left(\hat{\eta}\right)$
  - Sectoral productivities, $z\left(\hat{\eta}\right)$
  - Size of sectoral entry costs, $\left(I - \operatorname{diag}\left(\hat{\eta}\right)\right)\log\kappa$

Equilibrium is unique and efficient

III. Equilibrium returns to scale

RTS and sectoral productivity

Proposition

High average sectoral productivity leads to higher RTS:

$$\frac{d\hat{\eta}_i}{d\mu_j} = \left[(1-\varphi_i)\,\frac{d^2 a_i}{d\hat{\eta}_i^2}\right]^{-1} \mathcal{K}_{ij} \geq 0,$$

where $\mathcal{K} = \alpha\mathcal{L}$ is a matrix with nonnegative elements.

- Two effects of high μ_j
  1. PE: High μ_j means that sector-j firms become more productive $\Rightarrow \hat{\eta}_j \uparrow$
  2. GE: Due to competition, P_j goes down $\Rightarrow \hat{\eta}_j \downarrow$
- (2) benefits j's direct and indirect customers (firms with $\mathcal{K}_{ij} = \alpha_i^\top \mathcal{L}_{\cdot j} > 0$) $\Rightarrow \hat{\eta}_i \uparrow$
- Without supply chain links, $\alpha = 0$, (1) and (2) exactly offset each other
  - If $\alpha = 0$, the only factor of production is labor, which is in fixed supply

RTS and entry costs

Proposition

The impact of entry costs on RTS is given by

$$\frac{d\hat{\eta}_i}{d\log\kappa_j} = \left[(1-\varphi_i)\,\frac{d^2 a_i}{d\hat{\eta}_i^2}\right]^{-1}(1_{i=j} - \mathcal{K}_{ij}),$$

such that $d\hat{\eta}_i/d\log\kappa_j < 0$ if $i \neq j$.

- High entry cost $\kappa_j \Rightarrow P_j$ has to increase
- Higher P_j makes firms in j bigger, $\hat{\eta}_j \uparrow$, but j's customers smaller, $\hat{\eta}_i \downarrow$
  - If j is a strong indirect customer of its own goods, $\mathcal{K}_{ii}$, $\hat{\eta}_j$ can go down

RTS and effective productivity dispersion

Proposition

An increase in j's effective productivity dispersion $\varphi_j = \sigma_j^2/2\gamma_j$ ▸ Expression

1. Leads to higher RTS in other sectors, $\frac{d\hat{\eta}_i}{d\varphi_j} > 0$;
2. *May* lead to a decline in RTS in sector j.

- High φ_j means that j features more high-productivity firms $\Rightarrow P_j$ goes down
- Customers of j benefit $\Rightarrow \hat{\eta}_i \uparrow$
- Part (2): Positive link between z_j and φ_j is strongest under lower $\hat{\eta}_j$

Extension: Transportation costs

- We can easily extend the model to handle transportation costs

  Empirics

  - To use one unit of j in production, firms in i have to purchase $1+\tau_{ij}$ units of j

- Transportation costs reduce sectoral productivities

$$z_i(\hat{\eta}_i) = \mu_i - \hat{\eta}_i \log T_i - (1-\varphi_i)\, a_i(\hat{\eta}_i) - (1-\hat{\eta}_i)\log\sqrt{1-\varphi_i}$$

  - Here $T_i = \prod_{j=1}^{N}(1+\tau_{ij})^{\alpha_{ij}}$

Proposition

An increase in transportation cost τ_{ij} leads to a decrease in RTS $\hat{\eta}_i$.

- High τ_{ij} means that firms in i want to rely less on $j \Rightarrow \hat{\eta}_i \downarrow$
- Price of j goes down as a result $\Rightarrow \hat{\eta}_j \downarrow$

IV. Aggregate implications

Changes in the environment and GDP

Proposition

GDP increases in response to

1. Higher average sectoral productivity μ_j (Hulten);
2. Lower entry cost κ_j;
3. Higher effective productivity dispersion φ_j;
4. Lower transportation cost τ_{ij}.

- GDP is maximized in the efficient equilibrium $\Rightarrow$ Envelope theorem
- This is local result; for large changes in parameters, endogenous RTS
  - magnify changes that are beneficial for GDP
  - dampen changes that are harmful for GDP

GDP and RTS dispersion

Proposition

Consider an alternative economy in which all firms pick $\eta_{il} = \hat{\eta}_i$ irrespective of their ε_{il}. Then

$$y - y^{alt} = -\sum_{i=1}^{N} \omega_i (1 - \hat{\eta}_i) \log \sqrt{1 - \varphi_i} > 0.$$

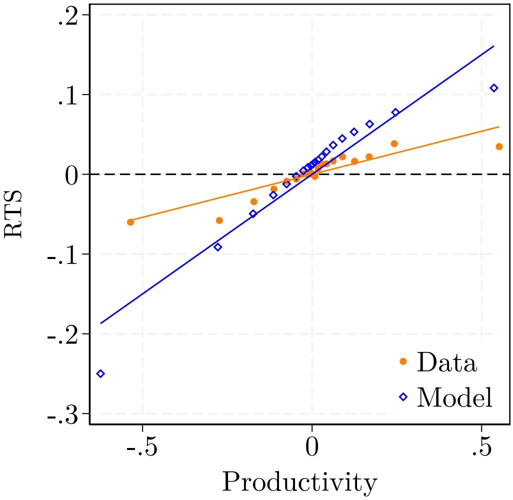
- In the two economies, sectors have the same sizes (Domar weights)
- But in the alternative economy, the economy is less productive
  - Productive firms cannot scale up their technologies

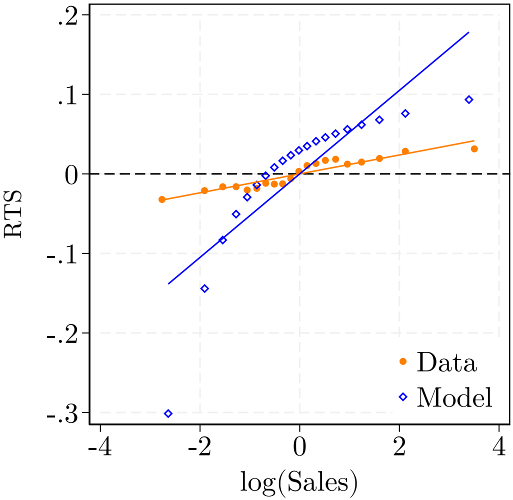
V. Quantitative exploration (preliminary!)

Calibration

- We calibrate the model to Spanish economy (62 sectors)
- Some parameters have direct empirical counterparts
  - **Household**: Consumption shares, β
  - **Firms**: Supply chain structure α, entry costs κ
- Left to choose: shock parameters, μ and σ; cost function parameter γ
  - σ and γ govern within-sector heterogeneity
    - Target sectoral interquartile range in log profits and RTS (Bloom et al., 2018) Calibration details
  - For given σ and γ, pick μ to match effective returns to scale $\hat{\eta}$ RTS estimation

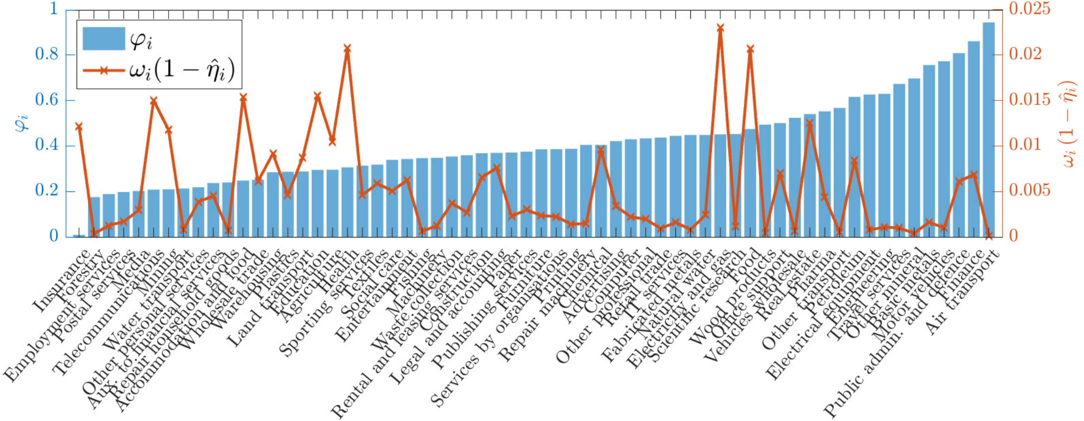
Productivity, size, and returns to scale: Data vs model





GDP and RTS dispersion

- Contribution of RTS dispersion to log GDP is $-\sum_{i=1}^{N} \omega_i (1-\hat{\eta}_i) \log \sqrt{1-\varphi_i} \approx 0.082$
  - Domar weights ω and effective RTS $\hat{\eta}$ are observable
  - $\varphi_i = \sigma_i^2 / 2\gamma_i$ is estimated to match within-sector IQR in RTS and profits



Conclusion

- A multisector model with endogenous RTS
  - Tractable aggregation
  - Matches key empirical facts
  - Endogenous RTS has a substantial aggregate effect
- Future research
  - Implications for growth

Appendix

Expression for ζ_{il}

- Normalization term ζ_{il} is

$$z_{il} = \left[(1-\eta_{il})^{1-\eta_{il}} \left(\left(1 - \sum_{j=1}^{N} \alpha_{ij} \right) \eta_{il} \right)^{\left(1-\sum_{j=1}^{N} \alpha_{ij}\right)\eta_{il}} \prod_{j=1}^{N} (\alpha_{ij}\eta_{il})^{\alpha_{ij}\eta_{il}} \right]^{-1}$$

- This functional form allows for a simple expression for marginal cost K

Back

Expression for Π_{il}

- We can write firm profits as

$$\log \Pi_{il} = \frac{1}{1-\eta_{il}} \left(\varepsilon_{il} - a_i \left(\eta_{il}\right) + \log P_i - \eta_{il} \log \left(W^{1-\sum_{j=1}^{N} \alpha_{ij}} \prod_{j=1}^{N} P_j^{\alpha_{ij}} \right) \right)$$

Back

Definition of Tornqvist index

Definition

Consider two firms, k and l, in sector i. Define relative productivity measures $\mathcal{T}_{ik}$ and $\mathcal{T}_{il}$ as $Q_{ik}/\mathcal{T}_{ik} = F_i\left(L_{ik}, X_{ik}, \eta_{il}\right)$ and $Q_{il}\mathcal{T}_{il} = F_i\left(L_{il}, X_{il}, \eta_{ik}\right)$. Then the base-firm invariant Tornqvist index is $\log_{\mathcal{T}} i, kl := \frac{1}{2}\left(\log \mathcal{T}_{ik} + \log \mathcal{T}_{il}\right)$, such that firm k is more productive than firm l is $\log \mathcal{T}_{i,kl} > 0$.

Back

Equilibrium uniqueness and efficiency

Proposition

There exists a unique equilibrium, and it is efficient. Equilibrium returns to scale vector $\hat{\eta}$ maximizes GDP.

◂ Back

RTS and effective productivity dispersion

Proposition

Consider $\chi_j \in \left\{\gamma_j, \sigma_j^2\right\}$. A change in χ_j leads to

$$\frac{d\hat{\eta}_i}{d\chi_j} = \left[(1-\varphi_i)\frac{d^2 a_i}{d\hat{\eta}_i^2}\right]^{-1}\left(\mathcal{K}_{ij}\frac{\partial z_j}{\partial \chi_j} + 1_{i=j}\frac{\partial^2 z_j}{\partial \chi_j \partial \hat{\eta}_j}\right),$$

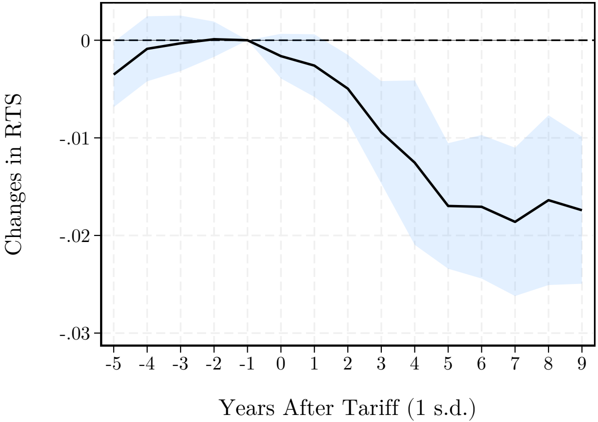
where

1. $\frac{\partial z_i}{\partial \sigma_i^2} = \left(a_j\left(\hat{\eta}_j\right) + \frac{1}{2}\frac{1-\hat{\eta}_j}{1-\varphi_j}\right)\frac{1}{2\gamma_j} > 0$
2. $\frac{\partial z_i}{\partial \gamma_i} = -\left(\frac{1}{\varphi_i}a_j\left(\hat{\eta}_j\right) + \frac{1}{2}\frac{1-\hat{\eta}_j}{1-\varphi_j}\right)\frac{\sigma_j^2}{2\gamma_j^2} < 0$
3. $\frac{\partial^2 z_i}{\partial \sigma_i^2 \partial \hat{\eta}_i} = \left(\frac{da_i}{d\hat{\eta}_i} - \frac{1}{2}\frac{1}{1-\varphi_j}\right)\frac{1}{2\gamma_i}$
4. $\frac{\partial^2 z_i}{\partial \gamma_i \partial \hat{\eta}_i} = -\left(\frac{1}{\varphi_i}\frac{da_i}{d\hat{\eta}_i} - \frac{1}{2}\frac{1}{1-\varphi_j}\right)\frac{\sigma_j^2}{2\gamma_i^2}$

In particular, $\frac{d\hat{\eta}_i}{d\sigma_j^2} \geq 0$ and $\frac{d\hat{\eta}_i}{d\gamma_j} \leq 0$ if $i \neq j$.

Tariffs and RTS

- Sector-year input tariff: $\log T_{it} = \sum_{c,j} \left(\frac{\text{Import Expenditure}_{t-1}^{(\text{Spain},i)\leftarrow(c,j)}}{\text{Total Material Expenses}_{i,t-1}} \times \log\left(1 + \text{Tariff}_t^{\text{Spain},(c,j)}\right)\right)$
  - Cost shares from OECD multi-country I-O table, tariff rate from Global Tariff Project (Teti, 2024)
- Panel local projection for horizons h: $RTS_{ilt+h} - RTS_{ilt-1} = \beta_h \log T_{it} + \gamma_{lh} + \gamma_{th} + \varepsilon_{ilth}$



RTS estimation

- For each year t, firms within size-percentile p_t in a sector i have the same production technology

$$Y_{ilt} = A_{ilt} K_{ilt}^{\beta_k^{i,p_t}} L_{ilt}^{\beta_l^{i,p_t}} M_{ilt}^{\beta_m^{i,p_t}}$$

- For each (i, p_t) cell, we apply the Blundell-Bond (2000) IV-GMM estimator for the model

$$\begin{aligned} y_{ilt} =& \beta_l^{i,p_t} n_{ilt} + \beta_l^{i,p_t} \rho^{i,p_t} l_{ilt-1} + \beta_k^{i,p_t} k_{ilt} + \beta_k^{i,p_t} \rho^{i,p_t} k_{ilt-1} + \beta_m^{i,p_t} m_{ilt} + \beta_m^{i,p_t} \rho^{i,p_t} m_{ilt-1} + \\ & \rho^{i,p_t} y_{ilt-1} + \gamma_t^{i,p_t} + \eta_i^{i,p_t} + \vartheta_{ilt}^{i,p_t} \end{aligned}$$

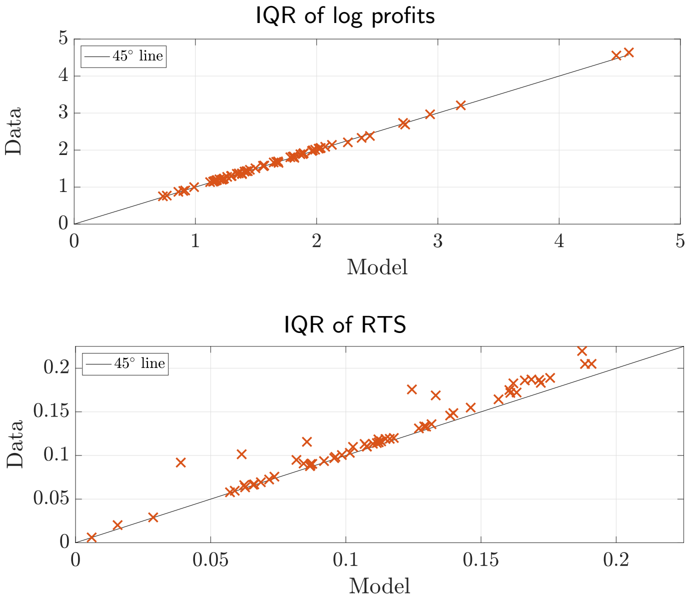
  on a rolling-window $(t-3$ to $t+3)$ rolling-percentile $(p-10$ to $p+10)$ sample in each sector

- Estimated RTS is

$$\eta_{ilt} = \beta_l^{i(l),p_t(l)} + \beta_k^{i(l),p_t(l)} + \beta_m^{i(l),p_t(l)}$$

Back

Goodness of fit



Back