Increasing Returns to Scale and Markups

Olga Shanks

George Mason University

ostaradu@gmu.edu

November 15, 2023
Motivation

• Increasing Returns to Scale can explain
  • Rising industry concentration
  • Decreasing share of labor in total output
  • Rising markups

Yet economists use the assumption of constant returns
• Autor et al. (2020) use CRS and require a change in consumer price sensitivity
• Karabarbounis and Neiman (2014) use CRS and require the capital-labor elasticity of substitution to be greater than one
• De Loecker et al. (2020) argue that markups cause industry concentration

• Estimation of Markups
  • De Loecker et al. (2020) argue that the aggregate markup of U.S. firms rose from 1.2 to 1.6 since 1980 to 2016
  • Inconsistent with profitability trends
  • Treatment of variable and fixed costs
  • Long vs. short horizons
Data

- Compustat Fundamentals Annual database
- Publicly traded companies in the U.S. from 1980 to 2019

Figure: Number of firms in Compustat by year by industry
Data (cont.)

- Variable costs: Cost of Goods Sold (COGS) + Selling, General and Administrative (SG&A)
- Capital costs: Property, Plant and Equipment (PPE) \times user cost of capital
- User cost of capital is estimated: \( r_t = i_t - \pi_t + \delta_t \)
  - \( i_t \): the Federal Funds rate
  - \( \pi_t \): FRED reported inflation rate, and
  - \( \delta_t = 12\% \) for depreciation and risk premium
- Revenues and costs deflated by BEA chain-type price indexes by industry (2- or 3-digit NAICS level)
- Excluded Finance sector (NAICS code 52)
- 5-year rolling periods, e.g. 1980-1984, 1981-1985, etc.
Results: Aggregate Returns to Scale

- Elasticity of scale is above 1
- Divergence after the Internet revolution
- OLS and Syverson’s methods are biased
- Focus on ACF because:
  - Most conservative estimate
  - Allows estimation of standard errors
  - Variable costs are dynamic like capital
Results: Industry-specific Elasticity of Scale

- Highest elasticity of scale
  - Manufacturing of wood and chemicals
  - Mining
  - Information

- Lowest elasticity of scale
  - Retail
  - Accommodation and Food
  - Transportation

- High volatility in the bottom 7 industries, likely due to the small number of observations
Markup Estimation

• Cost-minimizing firm:
\[ L(V, K, \lambda) = P^V V + rK - \lambda(Q(\Omega, V, K) - \bar{Q}) \]

• Derive from FOCs:
\[ \mu = e_V \frac{PQ}{P^V V}, \]  
(1)

where \( \mu \) is markup and \( e_V \) is output elasticity of the variable input

• According to Varian (1992), Syverson (2019) and others,
\[ \mu = MC \frac{P}{MC} \frac{AC}{AC} \frac{Q}{Q} = AC \frac{MC}{AC} \times Q = e_{scale} \frac{PQ}{TC} \]  
(2)

• De Loecker et al. (2020) use \( e_V \): \[ \mu = e_{COGS} \frac{Sales}{COGS} \]

Traina (2018) uses \( e_V \): \[ \mu = e_{COGS+SG&A} \frac{Sales}{COGS+SG&A} \]

Present research uses \( e_{scale} \): \[ \mu = e_{COGS+SG&A+capex} \frac{Sales}{COGS+SG&A+capex} \]
Variable Costs

Variable costs: COGS vs. COGS + SG&A?

- COGS typically include:
  - raw materials
  - direct labor
  - manufacturing overhead
  - freight in

- SG&A typically include:
  - wages of sales and office staff
  - shipping of finished goods
  - rent & utilities
  - R&D

- COGS have been going down, while SG&A have been trending up: firms have been shifting costs from COGS to SG&A

- Firms have the incentive to improve Gross Margin (i.e. Revenues - COGS)

- Firms in the same industry may ”decide” whether to record certain costs as COGS vs. SG&A

- Based on above: Variable costs = COGS + SG&A

  Traina (2018)’s $\mu = 1.2$ vs. De Loecker et al. (2020)’s $\mu = 1.6$
Choice of Markup Formula

Markup formula: \( \mu = e_v \frac{PQ}{PV^V} \) or \( \mu = e_{scale} \frac{PQ}{TC} \)?

- \( \mu = e_v \frac{PQ}{PV^V} \) reflects a short-term view, where firms cannot adjust capital
- \( \mu = e_{scale} \frac{PQ}{TC} \) reflects a long-term view, where all costs can change
- Long-term view is more appropriate for looking at data from 1980 to 2019
- Long-term view is more appropriate for large firms (most publicly traded firms)
- Hardware refresh cycles have been shrinking from 10 to 5 to 3 years with fast-changing technology
- Renting vs. owning real estate will result in different classifications of costs for firms in the same industry (SG&A for renting and capital for owning)
- Based on above: \( \mu = e_{scale} \frac{PQ}{TC} \)
Results: Decomposition of the Aggregate Markup

Figure: Decomposition of the Aggregate Markup into Elasticity of Scale and Sales-to-cost Ratio.
Results: Industry-specific Markups

- Highest markups
  - Information
  - Manufacturing of wood and chemicals
  - Manufacturing of food and textiles

- Lowest elasticity of scale
  - Utilities
  - Transportation
  - Construction

- High volatility in the bottom 7 industries, likely due to the small number of observations
Macroeconomic Implications

Constant Returns to Scale
\[ Y_i = z_i K^\alpha L^{1-\alpha} \]
Firms with higher \( z_i \) are more productive and get bigger

**Implications:** break up a big firm \( \Rightarrow \) same high \( z_i \) across many small firms \( \Rightarrow \) increased competition and efficiency

Increasing Returns to Scale
\[ Y_i = z_i L^\alpha K^\beta, \quad \alpha + \beta > 1 \]
Firms of bigger size are more productive

**Implications:** break up a big firm \( \Rightarrow \) same high \( z_i \) across many small firms \( \Rightarrow \) increased competition, but *destroys* productivity
Appendix
Estimation Methods: OLS and Syverson’s method

- **OLS**

\[ Y = AK^\alpha V^\beta \]  
\[ y_{it} = a + \alpha k_{it} + \beta v_{it} + u_{it}, \]

so \( \alpha + \beta \) measures the elasticity of scale

- **Syverson’s method**

\[ Y = A(K^\alpha V^{1-\alpha})^\gamma \]  
\[ y_{it} = a + \gamma \ln (K_{it}^\alpha V_{it}^{1-\alpha}), \]

where \( \alpha \) is the share of capital in total costs, and \( \gamma \) is the elasticity of scale

Note: all regressions are run on data within 5-year rolling periods and include year fixed effects and sub-industry fixed effects.
Issues with Estimation: Omitted Price Bias

- Simple OLS in logs as a starting point: \( y_{it} = \beta_0 + \beta_k k_{it} + \beta_v v_{it} + u_{it} \)
- Klette and Griliches (1996): output price is correlated with input choices
- Bond et al. (2021): deflating prices does not resolve the bias in the presence of market power and heterogeneous markups

\[
\begin{align*}
  r_{it} &= y_{it} + p_{it} = \beta_0 + \beta_k k_{it} + \beta_v v_{it} + p_{it} + u_{it},
\end{align*}
\]

where, in logs, \( r_{it} \) is revenue, \( y_{it} \) is output, \( p_{it} \) is price, \( k_{it} \) is capital, \( v_{it} \) is variable inputs, and \( u_{it} \) is the error term.

After deflating:

\[
\begin{align*}
  r_{it}^d &= \beta_0 + \beta_k k_{it} + \beta_v v_{it} + (p_{it} - p_{t\text{-index}}) + u_{it}
\end{align*}
\]

- Potential solution
  - add a proxy variable for \((p_{it} - p_{t\text{-index}})\)
  - share in total industry costs, \( s \)
  - \( s \) reflects relative firm size; size affects the firm’s residual demand, which in turn affects the price differential
Issues with Estimation: Simultaneity and Selection

\[ r_{it}^d = \beta_0 + \beta_k k_{it} + \beta_v v_{it} + s_{it} + u_{it} \]  \hspace{1cm} (9)

- Simultaneity
  - \( u_{it} \) contains productivity shock \( \Omega_{it} \)
  - Productivity shock affects inputs
  - Productivity shock is observed by the firm but unobserved by the econometrician

- Selection
  - Firms may respond to a negative productivity shock by exiting the market altogether

- Olley and Pakes (1996) and Ackerberg et al. (2015) resolve these biases
Estimation Methods: Olley-Pakes

\[
inv_{it} = inv_t(\Omega_{it}, k_{it}, s_{it}) \quad (10)
\]
\[
\Omega_{it} = h_t(inv_{it}, k_{it}, s_{it}) \quad (11)
\]
\[
 r_{it}^d = \beta_0 + \beta_v v_{it} + \beta_k k_{it} + \beta_s s_{it} + h_t(inv_{it}, k_{it}, s_{it}) + e_{it} \quad (12)
\]
\[
 \phi_{it} = \beta_k k_{it} + \beta_s s_{it} + h_t(inv_{it}, k_{it}, s_{it}) \quad (13)
\]

- estimate (6) with OLS using a second-order polynomial for \( \phi_{it} \)

\[
\Omega_{it} = g_t(\Omega_{it-1}, P_{it}) + \varepsilon_{it} \quad (14)
\]
\[
P_{it} = p_t(inv_{it-1}, k_{it-1}, s_{it-1}) \quad (15)
\]
\[
r_{it}^d - \hat{\beta}_v v_{it} = \beta_0 + \beta_k k_{it} + \beta_s s_{it} + g_t(\hat{\phi}_{it-1} - \beta_k k_{it-1} - \beta_s s_{it-1}, \hat{P}_{it}) + \varepsilon_{it} + e_{it} \quad (16)
\]

- estimate (9) with probit using a second-order polynomial for \( p_t \)
- estimate (10) with nonlinear least squares using a second-order polynomial for \( g_t \)
- all regressions are run on data within 5-year rolling periods and include year fixed effects and sub-industry fixed effects
Estimation Methods: Ackerberg-Caves-Frazer

\[ \Omega_{it} = h_t(\text{inv}_{it}, v_{it}, k_{it}, s_{it}) \]  \hspace{1cm} (17)

\[ r^d_{it} = \beta_0 + \beta_v v_{it} + \beta_k k_{it} + \beta_s s_{it} + h_t(\text{inv}_{it}, v_{it}, k_{it}, s_{it}) + e_{it} \]  \hspace{1cm} (18)

\[ \phi_{it} = \beta_v v_{it} + \beta_k k_{it} + \beta_s s_{it} + h_t(\text{inv}_{it}, v_{it}, k_{it}, s_{it}) \]  \hspace{1cm} (19)

- estimate (12) with OLS using a second-order polynomial for \( \phi_{it} \)

\[ E \left[ y_{it} - \beta_0 - \beta_v v_{it} - \beta_k k_{it} - \beta_s s_{it} + g_t(\hat{\phi}_{it-1} - \beta_v v_{it-1} - \beta_k k_{it-1} - \beta_s s_{it-1}, \hat{P}_{it}) \otimes \begin{pmatrix} v_{it} \\ k_{it} \\ s_{it-1} \\ \hat{P}_{it} \\ \hat{\phi}_{it-1} \end{pmatrix} \right] = 0 \]  \hspace{1cm} (20)

- estimate (14) with generalized method of moments using a second-order polynomial for \( g_t \)

- all regressions are run on data within 5-year rolling periods and include year fixed effects and sub-industry fixed effects
Results: Markups Using Different Cost Categories

Figure: Markups Using Different Cost Categories
References


