Data and Markups

Jan Eeckhout
Pompeu Fabra

Laura Veldkamp
Columbia

2024

Has Big Data Created Big Markups?

- Big firms benefit more from data (Brynjolfsson & co. '16, '20, '21).
  Not just big-5, but regular product firms too.

- Have markups risen? Literature is split:
  Market-wide markups suggest rising market power (counter-cyclical).
  Product level markups look stable (procyclical).
  De Loecker, Eeckhout & Unger '20, Philippon & Gutierrez '16, Furman & Orszag '15,
  Ramey & Nekarda '20, Bils '87

- Our question: How should data affect markups?
  Can data help us makes sense of this conflicting evidence?
  This is a model to inform measurement and make sense of existing facts.

Main Ideas / Outline

- Modelling data
  - Data is information. Information resolves uncertainty (risk).
    We model this as uncertain consumer demand for a variety of products.
  - Firms price risk.
  - Firms choose an up-front investment and then choose how much to produce.
- Data's Competing Effects on Markups: Cost vs. Risk
- Composition Effects
  - Firms use data to produce more profitable (high markup) goods.
    Data creates composition effects that grow with more data.
  - Procyclical product and counter-cyclical firm / industry markups
- Markup dynamics, data barter and firm scope

Model (1/2): Firms

- n_F firms, indexed by i.
- The product space has N attributes, indexed by j.
  Goods, indexed by k, are combinations of attributes with weights a_{jk}.
- Firm chooses: investment in lowering marginal cost $\tilde{\boldsymbol{c}}_i$ at price $g(\chi_c, \tilde{\boldsymbol{c}}_i)$, and a quantity to produce $\mathbf{q}_i$ (vector) to maximize $E[U_i]$ and U_i:

$$U_i = \mathbf{E}\left[\pi_i|\mathcal{I}_i\right] - \frac{\rho_i}{2}\mathbf{Var}\left[\pi_i|\mathcal{I}_i\right] - g(\chi_c, \tilde{\boldsymbol{c}}_i) \tag{1}$$

$$\pi_i = \tilde{\boldsymbol{q}}_i'\left(\tilde{\boldsymbol{p}} - \tilde{\boldsymbol{c}}_i\right) \tag{2}$$

- ρ_i is firm i's price of risk.

Model (2/2): Demand and Data

- A demand curve: Customers' willingness to pay decreases in the quantity that all firms produce of attribute j:

$$\tilde{\mathbf{v}}_{\mathbf{i}} = \underline{p} - \frac{1}{\phi} \sum_{i'=1}^{N} \tilde{\mathbf{q}}_i + \mathbf{b}_i \tag{3}$$

demand shock for attributes $\mathbf{b}_i \sim N(0, I)$, $corr(\mathbf{b}_i, \mathbf{b}_j) \epsilon \{0, I\}$.

- Customers' value of a good is linear in attributes (hedonic pricing) $v_{ik} = \sum_{j=1}^{N} a_{jk} \tilde{v}_{ij}$.

- Data is information about demand shocks:
  n_{di} data points for firm i (exogenous for now).
  Each data point is : $\tilde{\boldsymbol{s}}_{i,z} = \boldsymbol{b}_i + \tilde{\varepsilon}_{i,z}$, where $\tilde{\varepsilon}_{i,z} \sim N(\mathbf{0}, \Sigma)$.
  Information set: $\mathcal{I}_i := \{\tilde{\boldsymbol{s}}_{i,z}\}_{z=1}^{n_{di}}$ (data is private) or
  $\mathcal{I}_i := \{\{\tilde{\boldsymbol{s}}_{i,z}\}_{z=1}^{n_{di}}\}_{i=1}^{n_F}$ (public info)

Equilibrium

- FOC: Production depends on risk and price impact (denominator) and expected profit (numerator) Kyle '89 or Back and Zender '93.

$$\tilde{\boldsymbol{q}}_i = \boldsymbol{H}_i(\mathbf{E}[\tilde{\boldsymbol{p}}_i|\mathcal{I}_i] - \boldsymbol{c}_i) \qquad \text{where}$$

$$\boldsymbol{H}_i = \left(\rho_i \mathbf{Var}[\tilde{\boldsymbol{p}}_i|\mathcal{I}_i] + \frac{\partial \mathbf{E}[\tilde{\boldsymbol{p}}_i|\mathcal{I}_i]}{\partial \tilde{\boldsymbol{q}}_i}\right)^{-1}$$

Data lowers *Var*, raises $\mathbf{H}_i$.

$\mathbf{H}_i$ governs the $cov(q_i, p_i)$.

Data allows a firm to choose quantities that covary with prices.

- Optimal choice of cost (firm size):

$$\frac{\partial \mathbf{E}[U_i]}{\partial \tilde{\boldsymbol{c}}_{ij}} = \frac{1}{2} \underbrace{\frac{\partial \mathbf{E}[\tilde{\boldsymbol{p}} - \tilde{\boldsymbol{c}}_i]' \boldsymbol{H}_i \mathbf{E}[\tilde{\boldsymbol{p}} - \tilde{\boldsymbol{c}}_i]}{\partial \tilde{\boldsymbol{c}}_{ij}}}_{marginal\ \ benefit} - \underbrace{\frac{\partial g(\chi_c, \tilde{\boldsymbol{c}}_i)}{\partial \tilde{\boldsymbol{c}}_{ij}}}_{marginal\ \ cost} = 0 \ \ \forall j$$

Product Markups: Could Increase or Decrease

Product-level markup: $M^p_{ik} := \mathbf{E}[\boldsymbol{p}_i(k)]/\boldsymbol{c}_i(k)$.

Proposition 1

Data-investment complementarity. A firm with more data invests more (chooses a lower c_i).

Proposition 2

Lower costs, higher markups. More investment (lower c_i) in any attribute j s.t. $a_{jk} > 0$, increases the markup of good k.

Proposition 3

Data reduces markup risk premium Holding firm size fixed, an increase in about any attribute of good reduces its markup.

Don't look to markups to infer firms' data.

Aggregation Effects: Data and Firm Markups

- **Definition**: Firm-level markup is total revenue, divided by total cost

$$M_i^f := \frac{\mathbf{E}[\boldsymbol{q}_i'\boldsymbol{p}_i]}{\mathbf{E}[\boldsymbol{q}_i'\boldsymbol{c}_i]} = \frac{\mathbf{E}[\boldsymbol{q}_i]'\mathbf{E}[\boldsymbol{p}_i] + \mathbf{tr}\left[\mathbf{Cov}(\boldsymbol{p}_i, \boldsymbol{q}_i)\right]}{\mathbf{E}[\boldsymbol{q}_i'\boldsymbol{c}_i]}$$

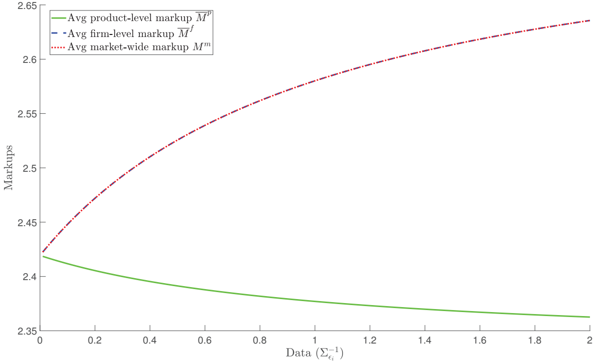
- Data increases $\mathbf{Cov}(\boldsymbol{p}_i, \boldsymbol{q}_i)$.
- Firms use data to create an aggregation effect: figure out which goods are more profitable and produce more of them.

Proposition 4

Data creates a wedge between product and firm markups. For high χ_c, an increase in firm i's data about any attribute increases $E[M_i^f - \overline{M}_i^p]$.

To Measure Data, Mind the Gap

Product and firm markups can both fall, both rise, or split.



- Symmetric firms. Parameters: $c_1 = c_2 = 1$. $\underline{p} = 5$, $\rho_1 = \rho_2 = 1$, $\phi = 0.1$, $A = I$.

The firm-product markup gap is a measure of a firm's data.

Industry Markup Measures Diverge

Proposition 5

Growing data increases

1. the difference between cost-weighted and unweighted firm markups $E[M^c - \overline{M}^f]$,
   (b/c high-data/ high-markup firms produce more.)
2. the difference between sales weighted and cost-weighted markups $E[M^s - M^c]$;
   (b/c high-data/ high-markup firms have higher sales, relative to costs.)
3. the difference between the sales weighted and industry-aggregates markup $E[M^s - M^{ind}]$.
   (b/c cost-weighted and industry-aggregated are the same.)

Diverging Industry Markup Measures

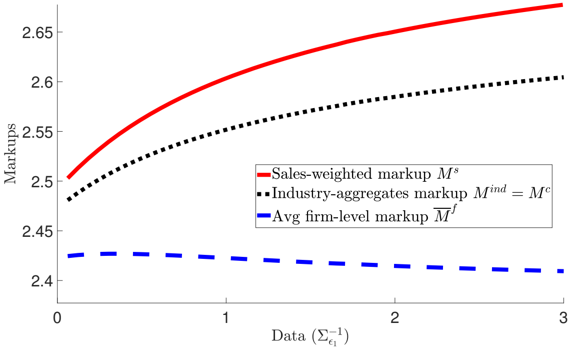
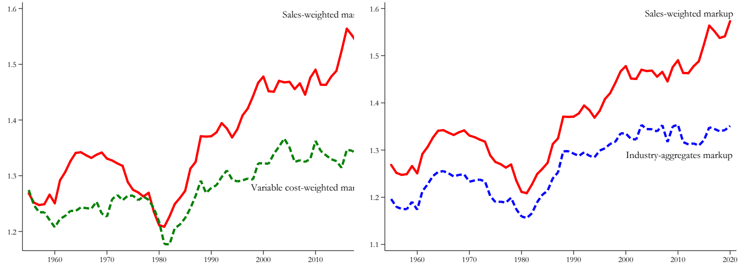


Figure: Data Accumulation Makes Industry Markup Measures Diverge. Investment cost function is $g(\chi_c, c_i) = \chi_c / c_i^2$, with $\chi_c = 1$. Parameters are $\bar{\rho} = 5$, $\rho_1 = 1$, $\rho_2 = 5$, $\phi = 0.8$ and $A = I$. Firm 1's data is measured on the x-axis. Firm 2's data is fixed at $\Sigma_{\epsilon_2}^{-1} = 1$.

Empirical Evidence: Firm/Industry Markup Divergence



2/3rds of the rise comes from sales-weighting large firms.

De Loecker, Eeckhout and Unger '20.

Data Can be Used for Product Innovation

- Let firm $i \in \{1, 2, \ldots, n_F\}$ choose an $n \times 1$ vector a_i that describes their location in the product space, such that $\sum_j a_{ij} = 1$.

- Firm's production problem is

$$max_{a_i, q_i} \mathbf{E}[\pi_i | \mathcal{I}_i] - \frac{\rho_i}{2} \mathbf{Var}[\pi_i | \mathcal{I}_i] - g(\chi_c, \tilde{\boldsymbol{c}}_i),$$

s.t. $\pi_i = q_i \boldsymbol{a}_i' (\tilde{\mathbf{p}} - \boldsymbol{c}_i)$ and $\sum_j a_{ij} = 1$.

- Result: This is just a linear rotation of the original problem. Data can inform what products I bring to market.

Dynamic Competition and Endogenous Data

- Each firm i chooses an $n \times 1$ vector a_{it} that describes their location in the product space and a quantity q_{it} to produce.

- Persistence of demand shocks b makes data a long-lived asset:

$$b_t = \rho b_{t-1} + \eta_{bt} \qquad \eta_{bt} \sim iid\ N(0, \sigma_\eta I) \tag{4}$$

  Transitory noise keeps all uncertainty from being resolved:

$$\tilde{b}_t = b_t + \epsilon_{bt} \qquad \epsilon_{bt} \sim iid\ N(0, \sigma_\epsilon I) \tag{5}$$

- Data is private and a by-product of economic activity: $n_{it} = q_{i,t-1} a_{i,t-1}$.
  Firms get more data about attributes they produce.
  $\rightarrow$ Active experimentation in the product space.

- Firms maximize present value of profit, V (Bellman eqn in data).

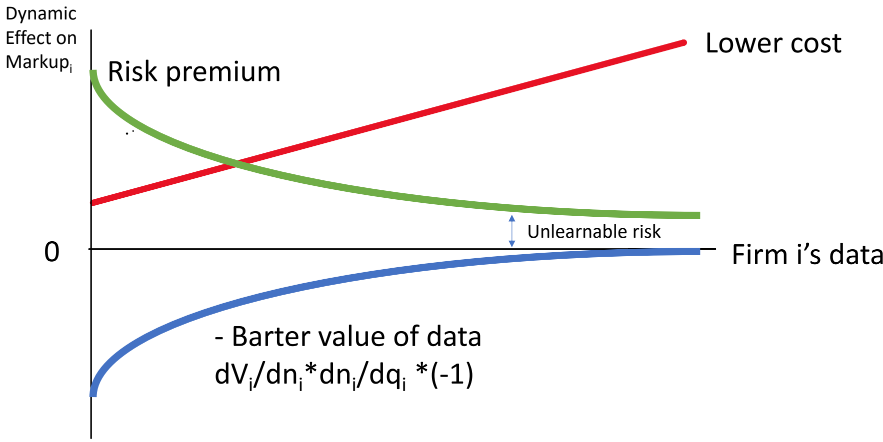
Production and Product Markups (dynamic)

- FOC: Production depends on risk and price impact (denominator) and expected profit, including data value (numerator)

$$\boldsymbol{q}_i a_i = \left(\rho_i \mathbf{Var}\left[\tilde{\boldsymbol{p}}|\mathcal{I}_i\right] + \frac{\partial \mathbf{E}\left[\tilde{\boldsymbol{p}}|\mathcal{I}_i\right]}{\partial \boldsymbol{q}_i}\right)^{-1} (\mathbf{E}\left[\tilde{\boldsymbol{p}}|\mathcal{I}_i\right] - \boldsymbol{c}_i + \frac{\partial V}{\partial q_i})$$

- New force in dynamic model: $\partial V/\partial q_i$ is the value of data.
  Payment p and **data barter** are substitutes.

Lifecycle of Firm Markups



As data firms mature and have abundant data, your data is worth less to them.

They raise prices.

Conclusions

- Start from a simple premise: Firms use data to predict what consumers want. Firms prefer higher profits and lower risk.
- Look for data in the covariances, not the levels of markups.
- Markups are higher at higher levels of aggregation.
  Firm and industry markups are more counter-cyclical.
  This reconciles empirical evidence.
- Dynamics: Where should a firm locate in product space?
  - Diversity or specialize?
  - A firm lifecycle theory of firm scope.