Market Power - Measurement

Winter 2024

Market Power and Productivity

- secular trend: the emergence of “superstar firms”
- does this reflect productivity or an increase in market power?
- what are the consequences for inequality, welfare, policy... ?
- This slide deck: Measurement of markups and productivity
- Next: Theories with endogenous competition and market power

The firm size distribution

- very robust finding: firm size distribution has a long upper tail
- this holds within the majority of industries, countries and after conditioning on observables
- typically, the size distribution is approximated with a lognormal or Pareto distribution
- broader theme: firms exhibit tremendous heterogeneity with respect to almost any variable we look at

Gibrat's law

- Gibrat's law states that if the growth rate of a variable is independent of its size, it will have a log-normal distribution in the long-run
- bare-bone model: let Y_{it} denote firm i's size in year t. Suppose it evolves according to:

$$Y_{it}/Y_{i,t-1} - 1 = \varepsilon_{it},$$

  where ε_{it} is i.i.d. across firms and time
- then, after allowing a large group of firms to evolve a while, the cross-section of firm sizes will have a log-normal distribution

Log-normality in Portugal: Cabral Mata 2003

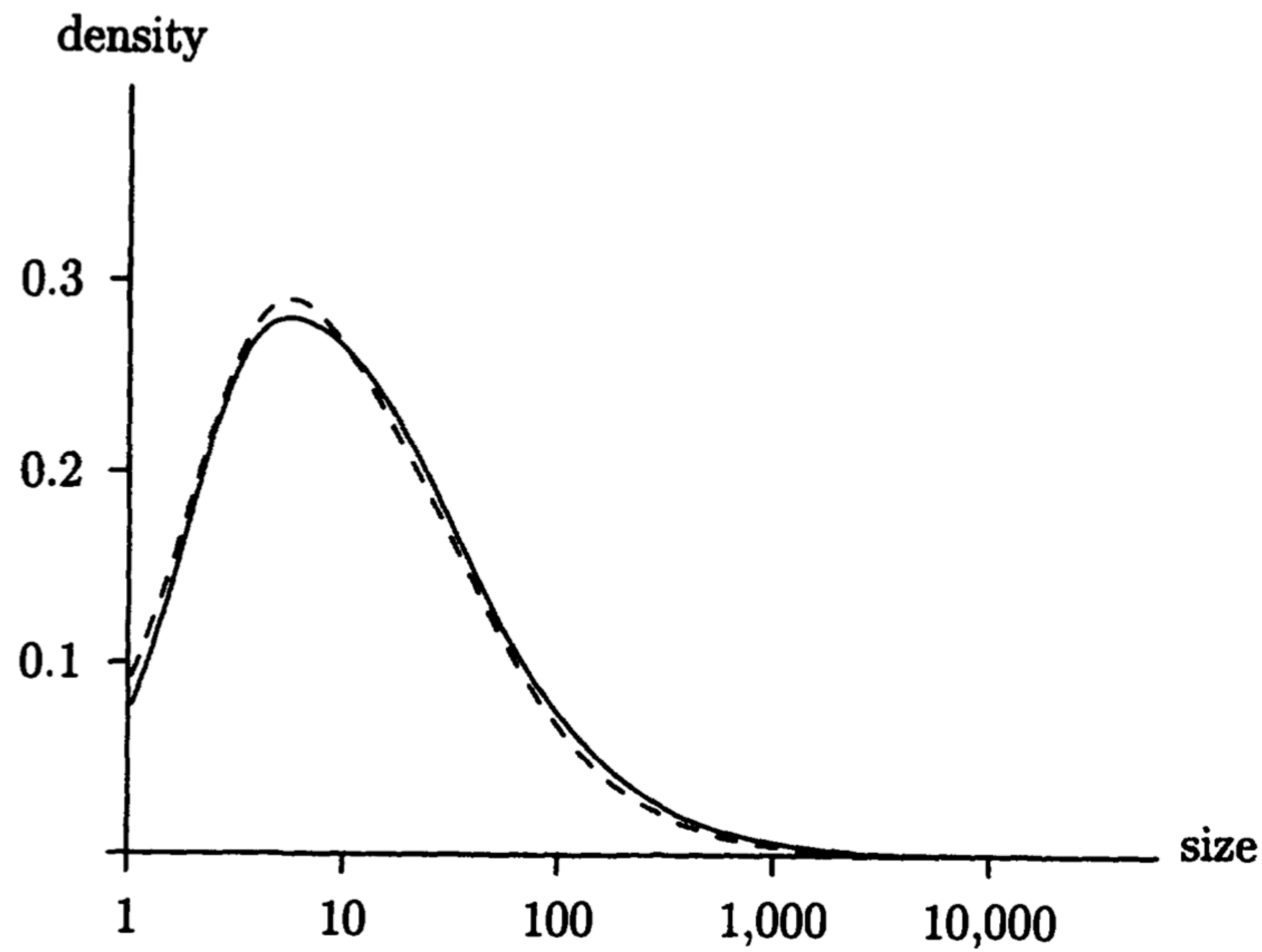
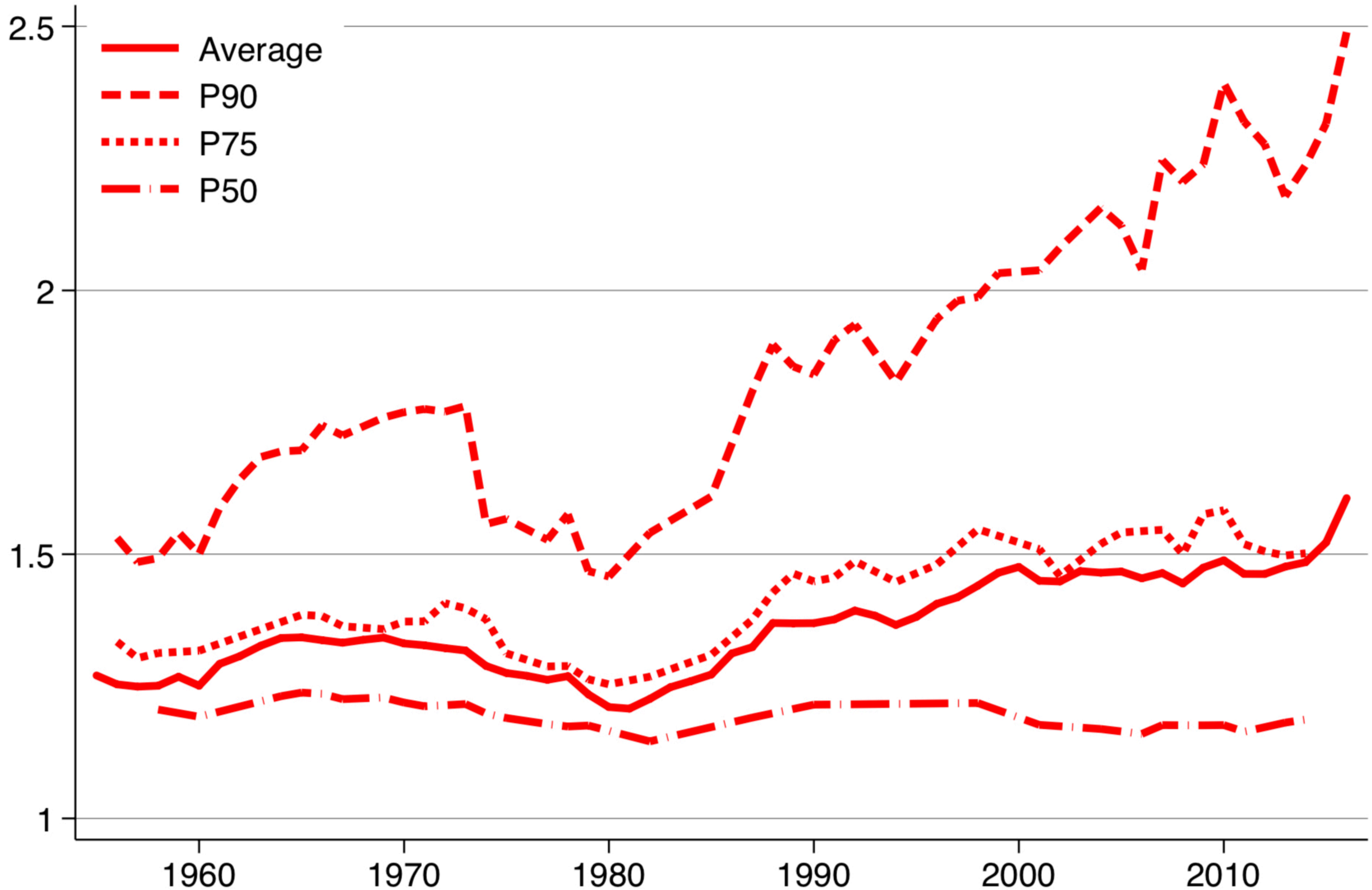


FIGURE 2. FIRM SIZE DISTRIBUTION IN 1983 (SOLID LINE) AND 1991 (DASHED LINE), BASED ON EMPLOYMENT DATA FROM THE *QUADROS DO PESSOAL* DATA SET

Theoretical models of industry dynamics

- “Gibrat model” is too simple to explain the data, but it is in the background: most modern models of heterogeneous firms are based on the assumption that firms experience random, time-varying shocks
- Jovanovic 1982: entry and exit model that explains systematic relationship between variable growth rates, exit rates and firm size
- Hopenhayn 1992: equilibrium model with stochastic productivity
- Melitz 2003: equilibrium model with selection intro trade
- in all of the above, changes in firm size distribution are interpreted as changes in the underlying productivity distribution
- reason: exogenous market power, i.e., markups

Markup variation over time



De Locker, Eeckhout, Unger (2020)

Markups vs Productivity

- market power is a source of misallocation
- if markups are endogenous, changes in firm dynamics could reflect market power or productivity
- posts a challenge for theory and measurement alike
- yet, many secular trends are likely intertwined with market power dynamics
  - Falling labor share, declining business dynamism, wage stagnation, secular trends in interest rates
- lots of interesting work to be done - theoretically and empirically

Estimating Market Power

- literature has focused on price markups

$$p_i = \mu_i \times mc_i = \frac{\sigma_i}{\sigma_i - 1} \times mc_i$$

- to main approaches:
    1. demand-based methods: estimate the residual demand curve
    2. production-based methods: estimate production function
- will mostly focus on 2.: main focus of macro/trade/spatial literature; forces us to deal with markups vs productivity
- papers that compare 1. and 2. in the same setting are rare
    - See de Loecker and Scott (2016) for an exercise like that for the US beer industry

Method 1: Demand-based methods

- by far the most common approach in the field of IO
  - See e.g., Ackerberg et al (2007, Handbook chapter)
- basic idea is to imagine that within some industry grouping with J products we can estimate the demand system:

$$Q_i = d_i(\mathbf{P}), \quad \forall i$$

- Then assume some sort of "conduct", or market structure
  - i.e., the game that producers are playing in the model

Method 1

- Firm's FOC can be written as:

$$P_i = \mu_i MC_i \qquad \mu_i \equiv \frac{\sigma_i}{\sigma_i - 1}$$

- σ_i : perceived price elasticity of demand (residual elasticity) given by:

$$\sigma_i = -\frac{dQ_i}{dP_i} = -\left(\frac{\partial Q_i}{\partial P_i} + \sum_{j\neq i} \frac{\partial Q_i}{\partial P_j}\frac{dP_j}{dP_i}\right)$$

- special cases (e.g., perfect/monopolistic competition, Bertrand, Cournot, Collusion,...) restrict $\frac{dP_i}{dP_j}$

Method 1: Demand-based

- demand estimation is hard
  1. curse of dimensionality
  2. hard to find instruments
  3. which conduct to assume?
- little application outside of IO
  - data requirements are substantial, hard to go beyond particular industries/ product categories

Method 2: Production-based

- basic idea: Use firm production data (output and inputs) to effectively measure something like MC (and then just take $\mu = P/MC$)
- intellectual history of the current approach
  - Hall (JPE 1988)
  - Olley Pakes (1999), ...
  - De Loecker Warczynski (AER 2012)

de Loecker Warzynski

- idea: Start from cost-minimization problem of the firm
  - markups related to input cost shares and output elasticities
  - hard part is to estimate the output elasticity
- assumptions
  - firm has production function $Q_{it} = F(X_{it}^1, \ldots, X_{it}^V, K_{it}, \theta_{it})$ where X denotes variable inputs, K is capital and θ_{it} is productivity
  - variable input prices P_{it}^X are taken as given
  - firms minimize cost
  - no further restrictions on demand curve or conduct

de Loecker Warzynski: Sketch

- the Lagrangian for cost-minimization is given by

$$\mathscr{L}\left(X_{it}^1, \ldots, X_{it}^V, K_{it}, \lambda_{it}\right) = \sum_{v=1}^{V} P_{it}^v X_{it}^v + r_{it} K_{it} + \lambda_{it}(Q_{it} - Q(\cdot))$$

- first-order condition:

$$P_{it}^v = \lambda_{it} \frac{\partial Q_{it}(\cdot)}{\partial X_{it}^v}$$

  - λ_{it}= Lagrange Multiplier = marginal cost at Q_{it}

de Loecker Warzynski: Sketch

$$P_{it}^{v} = \lambda_{it} \frac{\partial Q_{it}(\cdot)}{\partial X_{it}^{v}}$$

- Multiply by X_{it}^{v}/Q_{it}

$$\frac{P_{it}^{v} X_{it}^{v}}{Q_{it}} \frac{1}{\lambda_{it}} = \frac{X_{it}^{v}}{Q_{it}^{v}} \frac{\partial Q_{it}(\cdot)}{\partial X_{it}^{v}}$$

- since $MC_{it} = \lambda_{it} = P_{it}/\mu_{it}$, can rewrite this as:

$$\frac{P_{it}^{v} X_{it}^{v}}{P_{it} Q_{it}} \mu_{it} = \frac{X_{it}^{v}}{Q_{it}^{v}} \frac{\partial Q_{it}(\cdot)}{\partial X_{it}^{v}}$$

The markup formula

- The leads to the simple expression

$$\mu_{it} = \theta_{it}^{v} / \alpha_{it}^{v}$$

  - θ_{it}^{v}: The output elasticity with respect to input v
  - α_{it}^{v}: The expenditure cost share of input v
- this is essentially Hall's insight: Whenever a variable input's output elasticity is greater than the input's revenue share, the difference is the markup $(\mu_{it} > 1)$ - but here no need to impose CRTS!
- implementation:
  - input shares are easily observed
  - to get the input's output elasticity, requires estimating productivity

Production function estimation

- consider a Cobb-Douglas technology (everything in logs)

$$q_{it} = \theta^l l_{it} + \theta^k k_{it} + \omega_{it} + \epsilon_{it}$$

  - hicks-neutral productivity $\exp(\omega_{it})$
  - ϵ_{it}: idiosyncratic productivity shock/measurement error
- productivity ω_{it} is observed by firm, not by econometrician
  - input choices respond to unobserved ω_{it}
  - OLS, therefore, suffers from endogeneity
- hence, recovering the production function is to estimate productivity

Productivity dispersion

- Bartelsman and Doms (2003) review some work on productivity
  - large productivity dispersion
  - within firm, productivity is highly but imperfectly persistent
  - there is considerable reallocation within industries
- De Loecker and Syverson (2021) report that 90-10 percentile TFP ratios of 2:1 are typical

What to make of these residuals?

- “I found the spectacle of economic models yielding large residuals rather uncomfortable, even when the issue was fudged by renaming them technical change and claiming credit for their measurement” - Zvi Griliches
- bad data could be one reason for TFP dispersion, but we observe large dispersion everywhere we have data, and measured productivities are connected to real outcomes
  - more productive firms are less likely to exit
  - more productive firms are more likely to export
  - entrants tend to have lower productivity than average incumbent

Thinking about bias

- how does simultaneity of input decisions bias the labor coefficient?
  - up: when productivity is high, a firm uses more labor
- selection due to exit can bias the capital coefficient estimate down
  - firms with high capital have lower exit cutoffs. thus, conditional on survival, there is a negative correlation between k and ω.
- another potential source of bias: measurement error, see Collard-Wexler and De Loecker (2016)

Olley-Pakes (1996)

- key idea: address the simultaneity problem by imposing additional structure on firms factor input decisions
- consider a firm that maximizes the present discounted value of current and future profits
- assume the observed productivity term ω_{it} evolves exogenously according to the Markov process

$$p(\omega_{it+1} \,|\, I_{it}) = p(\omega_{it+1} \,|\, \omega_{it})$$

- I_{it}: Information set at time t
- hence, $\omega_{it} = \mathbb{E}(\omega_{it} \,|\, \omega_{it-1}) + \xi_{it}$, where $\mathbb{E}(\xi_{it} \,|\, I_{it-1}) = 0$
- labor is assumed to be a static input chosen optimally each period with zero adjustment costs

Productivity inversion

- Assume that a firms optimal investment is a strictly increasing function of their current productivity ω_{it}

$$i_{it} = h_t\left(\omega_{it}, k_{it}\right)$$

- h_t captures input prices etc.
- given monotonicity, optimal investment can be inverted for productivity

$$\omega_{it} = h_t^{-1}(i_{it}, k_{it})$$

- this inverse function can be used to non-parametrically control for the productivity in the production function

OP first-stage

- Substitute inverse function into the production technology

$$q_{it} = \theta^l l_{it} + \theta^k k_{it} + h_t^{-1}(i_{it}, k_{it}) + \epsilon_{it}$$

- model the inverse function non-parametrically, which yields:

$$q_{it} = \theta^L l_{it} + \Phi_t(i_{it}, k_{it}) + \epsilon_{it}$$

- the coefficient on capital is not identified in the first-stage
  - colinear with the non-parametric function in i_{it} and k_{it}
- But can obtain estimates for Φ_t and θ^L in the first-stage, using a non-parametric regression
- Ackerberg, Craver and Frazer (2015) think more carefully about what is identifying θ^l, for now, we don't worry

First-stage output

- with $\hat{\theta}^l$, we can estimate $\Phi_t(i_{it}, k_{it})$:

$$\hat{\Phi}_t = q_{it} - \hat{\theta}^l l_{it}$$

- with these estimates, we would like to separate $\theta^k k$ and ω, which are both in the control function.
- we are going to use the Markov assumption on ω for identification

OP Second Stage

- Productivity process: $\omega_{it} = E(\omega_{it} | \omega_{it-1}) + \xi_{it}$
- ξ_{it} satisfies $E(\xi_{it} | I_{it-1}) = 0$
- Since $k_{it} \in I_{it-1}$, this implies: $E(\xi_{it} | k_{it}) = 0$
- Independence implies $E(k_{it}\xi_{it}) = 0$
- This supplies a moment condition to estimate θ^K
- GMM procedure:

1. Start with a guess for θ^K
2. Compute $\omega_{it}(\theta^K) = \hat{\Phi}_t - \theta^K k_{it}$
3. Compute ξ_{it} from regressing ω_{it} on ω_{it-1}
4. Compute the sample analog to the moment condition above

OP summary

- To identify the labor elasticity, use information in firms investment decisions to control for productivity shocks that is correlated with labor inputs
- Assume capital is determined before unobserved productivity realizes to estimate capital elasticity
- This approach can be implemented with more general technologies than Cobb-Douglas

Where it went from OP

- Investment can be lumpy and there are many zeros in the data
- Levinson and Petrin (2003): propose to use intermediate inputs m_{it}
  - Model it as an additional input under the same assumptions that it is strictly increasing in productivity
  - Then the first stage becomes: $q_{it} = \theta^L l_{it} + \Phi(m_{it}, k_{it}) + \epsilon_{it}$
- Ackerberg Caves Frazer 2015: argue that both OP and LP suffer from identification issues, at least in principle
- they propose a new approach which involves modified assumptions on the timing of input decision and removes identification of all coefficients of the production function to the second-stage

ACF approach

- Abandon the attempt to estimate the labor coefficient in the first stage
- Timing assumption: Labor is chosen after capital is chosen in the previous period, but before materials were chosen at time t
- Under these timing assumptions: $m_{it} = f_t(\omega_{it}, k_{it}, l_{it})$
- Under monotonicity, this can be estimated and used to substitute for ω_{it} in the production function

ACF procedure, used in De Loecker Warzyinski

- We can write:

$$q_{it} = \Phi_t(m_{it}, k_{it}, l_{it}) + \epsilon_{it}$$

$$\Phi_t(m_{it}, k_{it}, l_{it}) = \theta^k k_{it} + \theta^L l_{it} + f_t^{-1}(m_t, k_{it}, l_{it})$$

- First-stage estimates the control function Φ_t non-parametrically
- The moment condition to identify θ^k is the same
- Need an additional moment condition to identify labor
  - Note that $\xi_{it} = \omega_{it} - \omega_{it-1}$ is orthogonal to lagged labor inputs, since this is in the information set at $t - 1$. Essentially, lagged labor is an instrument for current labor.
  - So given a guess for both θ^k and θ^l, can compute $\omega_{it}(\theta^k, \theta^l) = \hat{\Phi}_{it} - \theta^k k_{it} - \theta^l l_{it}$
  - Then compute ξ_{it} as before, and compute the moment analogues to $\mathbb{E}(k_{it}\xi_{it}) = \mathbb{E}(l_{it-1}\xi_{it}) = 0$

Putting it together

- Individual mark-up

$$\mu_{it} = \theta^v_{it}(\alpha^v_{it})^{-1}$$

- Average mark-up (weighted by cost-share, sales, employment,...)

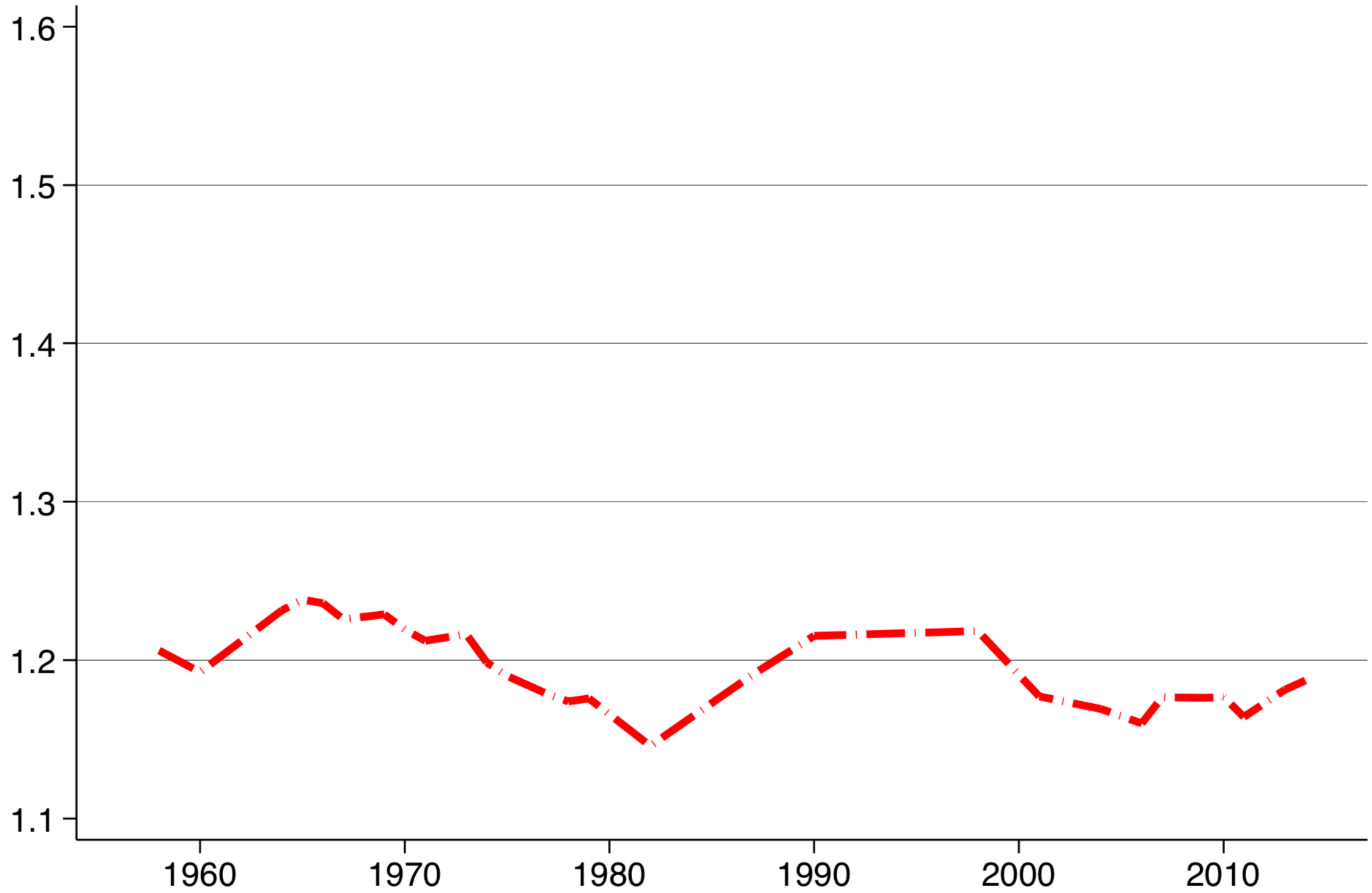
$$\mu_t = \sum_i m_{it}\mu_{it}$$

- Note: this methodology can also be applied to measure mark-down on wages so long as one factor v is competitively sourced:

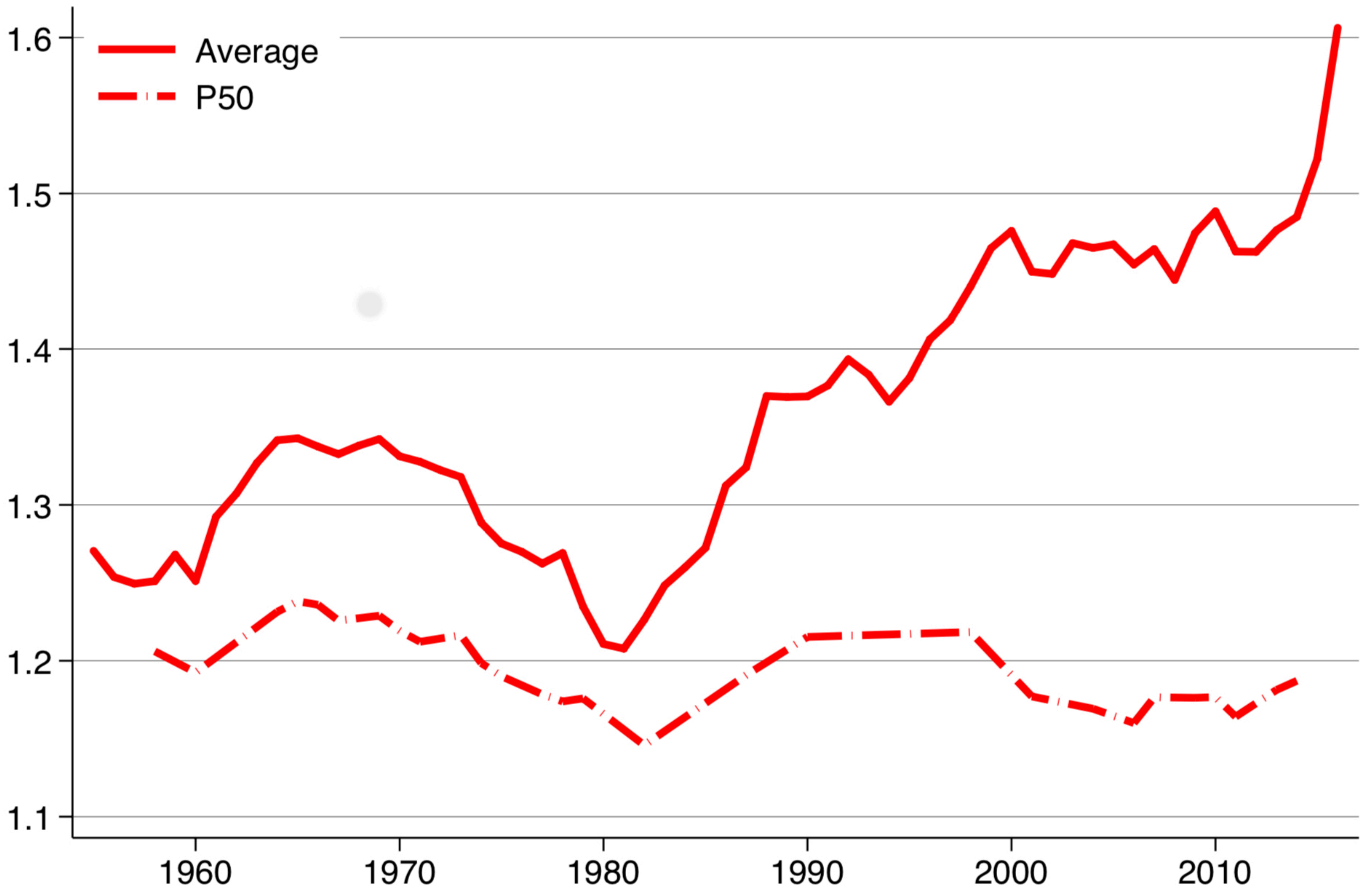
$$\frac{wL}{pQ} = \frac{\mathscr{M}}{\mu}\alpha^L$$

$$\frac{p^v X^v}{pQ} = \frac{1}{\mu}\alpha^v$$

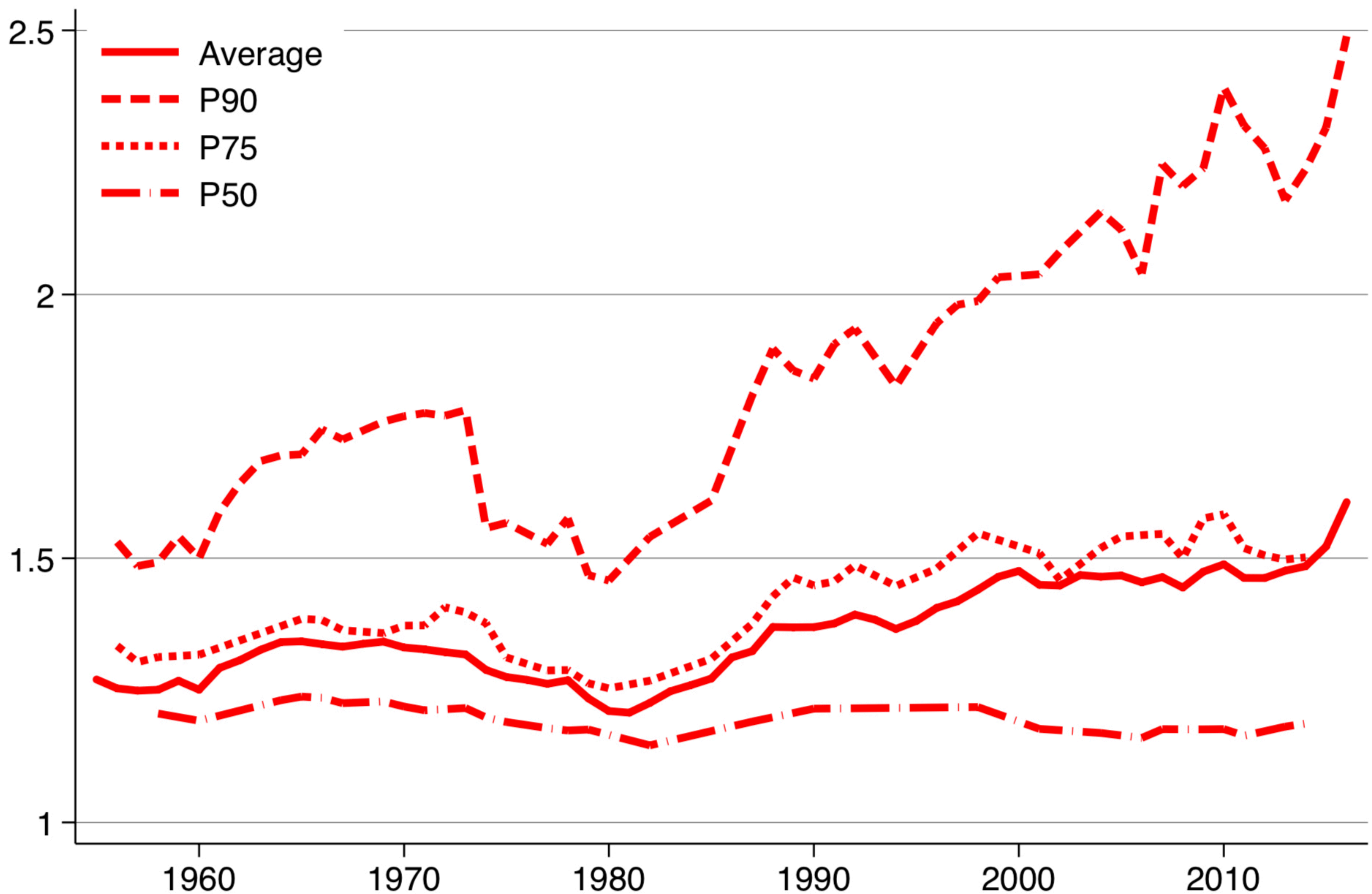
US Median Markup (De Loecker at al 2020)



Average Markups in the U.S.



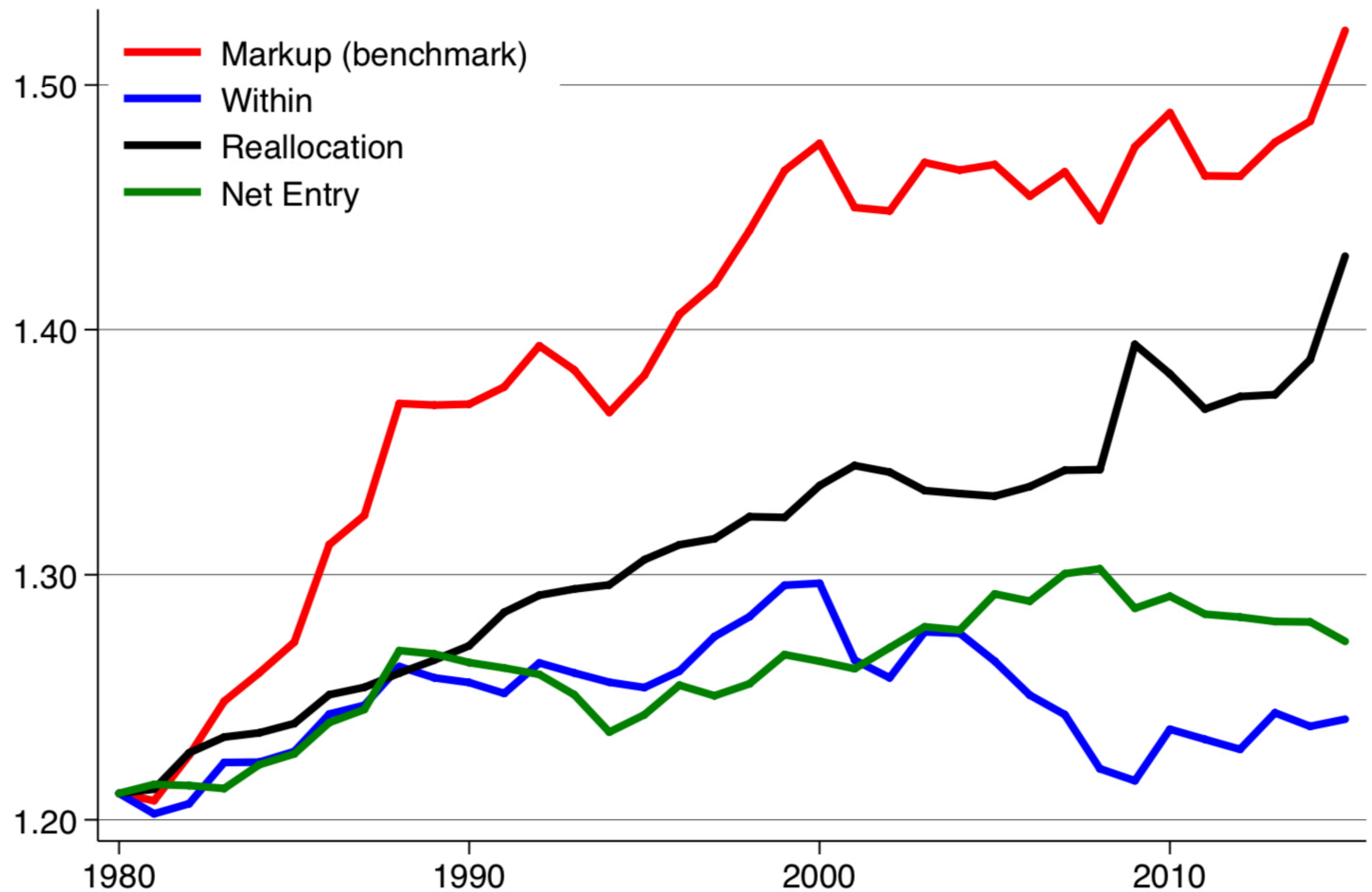
Changes in the upper tail drive aggregates



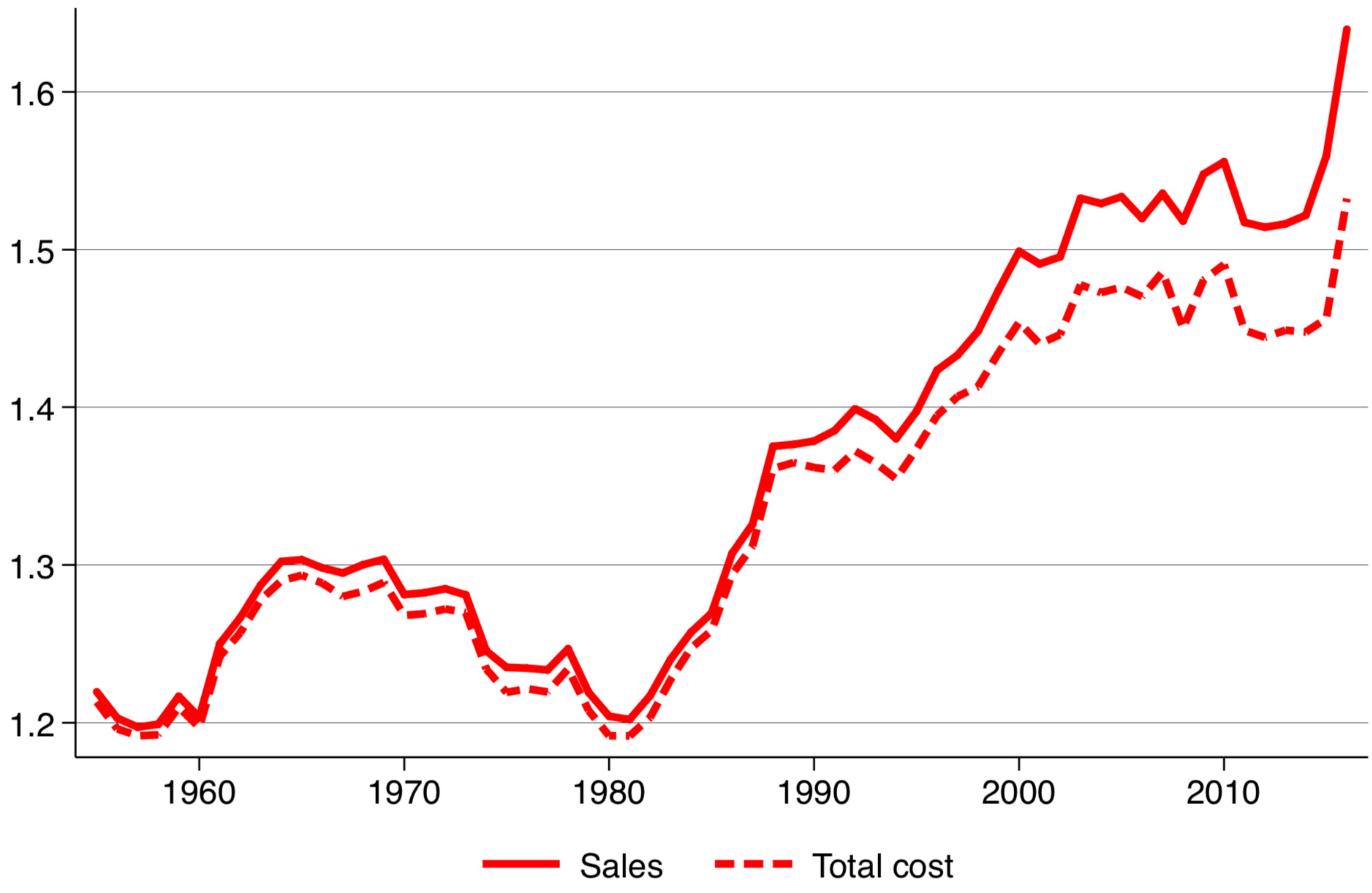
Reallocation

- Why are average markups rising?
  1. Within effects: Some firms have raised markups a lot
  2. Between effects: Reallocation towards high-markup firm

$$\Delta\mu_{it} = \sum_i m_{i,t-1}\Delta\mu_{it} + \sum_i \mu_{i,t-1}\Delta m_{it} + \text{Cross-Terms} + \text{Net entry}$$



Weights matter!



Weights matter