Chapter 22: Heterogeneous Firms

Chapter author: Toshihiko Mukoyama

April 2026

Outline

1. Introduction
2. A Simple Model
3. Firm Heterogeneity in the Data
4. Reallocation and Misallocation
5. Firm Heterogeneity in General Equilibrium
6. Alternative Market Arrangements
7. Business Cycles and Heterogeneous Firms
8. Endogenous Productivity: Klette-Kortum (2004)
9. Summary

Introduction

Motivation

Standard macroeconomic models assume an **aggregate production function**:

$$Y = F(K, L)$$

- Justified if all firms have homogeneous production functions
- In reality, firms are heterogeneous in many dimensions:
    - Large vs. small
    - Young vs. old
    - Growing vs. contracting
    - Productive vs. unproductive
    - Differentiated products

Why Does Firm Heterogeneity Matter?

Key policy and research questions that require explicit firm heterogeneity:

- How should we encourage (or discourage) the **entry of new firms**?
- Should we support **growing firms**, and if so how?
- Should we be concerned about the growing prominence of **"mega-firms"**?
- What are the causes and consequences of the recent **decline in business dynamism**?
- Does the representative-firm assumption give **misleading answers** to standard policy questions?

Chapter Roadmap

1. A simple model showing how heterogeneity affects aggregation
2. Empirical facts on U.S. firm heterogeneity
3. Reallocation and misallocation
4. General equilibrium: Hopenhayn-Rogerson (1993) framework
5. Alternative market arrangements (monopolistic competition, oligopoly)
6. Business cycles and heterogeneous firms
7. Endogenous productivity: Klette-Kortum (2004) model

A Simple Model

Setup

- Unit mass of firms, $i \in [0, 1]$
- Homogeneous good, perfect competition
- Production function of firm i:

$$y_i = a_i F(\mathbf{x}_i)^\gamma, \qquad \gamma \in (0, 1)$$

- a_i: firm-specific productivity (heterogeneous across firms)
- $\mathbf{x}_i$: input vector for firm i
- $F(\mathbf{x}_i)$: constant returns to scale
- Overall production: **decreasing returns to scale** (because $\gamma < 1$)

Why Decreasing Returns?

Key modeling assumption: $\gamma < 1$ ensures DRS at the firm level

With constant returns ($\gamma = 1$):

- The most productive firm(s) take over the entire economy
- Either: monopoly/oligopoly $\Rightarrow$ contradicts perfect competition
- Or: only the most efficient firms operate $\Rightarrow$ replicates homogeneous case

With $\gamma < 1$:

- Firms of different productivities can coexist
- Each firm has an optimal finite scale
- Richer implications for aggregation

Two-Step Optimization

Step 1: Cost minimization (common across firms)

$$\min_{\mathbf{x}} \; \mathbf{px} \quad \text{s.t.} \quad F(\mathbf{x}) = 1$$

Let $\mathbf{x}^*$ be the solution, $c \equiv \mathbf{px}^*$ the unit cost.

Step 2: Optimal scale for firm i

Let $m_i = F(\mathbf{x}_i)$ be combined input. Then:

$$\max_{m_i} \; a_i m_i^\gamma - c m_i$$

First-order condition:

$$a_i m_i^{\gamma-1} = \frac{c}{\gamma}$$

Aggregation

From the FOC: $y_i = (c/\gamma)m_i$ for all i. Summing:

$$Y = \frac{c}{\gamma}M$$

where $Y = \int y_i \, di$ is total output and $M = \int m_i \, di = F(\mathbf{X})$.

Define **aggregate productivity**:

$$A \equiv \left(\int a_i^{\frac{1}{1-\gamma}} \, di \right)^{1-\gamma}$$

Then the **aggregate production function** is:

$$Y = AF(\mathbf{X})^{\gamma}$$

Heterogeneity matters through A—the distribution of a_i shapes aggregate output.

Lognormal Example

Suppose $\log(a_i) \sim N(\nu - \sigma^2/2,\ \sigma^2)$

- Average productivity: $\int a_i \, di = \exp(\nu)$
- Aggregate productivity:

$$A = \exp\left(\nu + \frac{\gamma}{1-\gamma}\frac{1}{2}\sigma^2\right)$$

Key insight: Even holding the average of a_i fixed, a higher dispersion σ raises A.

Intuition: Productive resources are endogenously allocated—productive firms use more input and have a greater presence in aggregate production. More dispersion $\Rightarrow$ more room to allocate to productive firms in the right tail.

Two Fundamental Questions

Throughout this chapter, we keep two questions in mind:

1. **How is the distribution of a_i determined?**
    - Exogenous vs. endogenous productivity
    - Entry, exit, and innovation
2. **How does the economy allocate resources to different firms?**
    - Efficient allocation vs. misallocation
    - Frictions, distortions, and policy

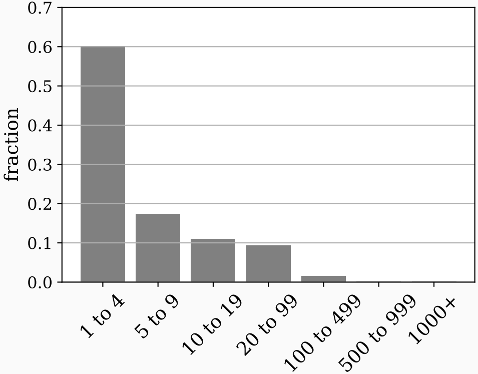
Firm Heterogeneity in the Data

Data Source

- All figures from the U.S. Census Bureau's **Business Dynamics Statistics (BDS)**
- Covers the universe of U.S. employer establishments and firms
- Key distinction:
    - **Establishment**: a fixed physical location of economic activity
    - **Firm**: a collection of establishments under common ownership
- Over 5 million firms and 7 million establishments in the U.S.

Firm Size Distribution

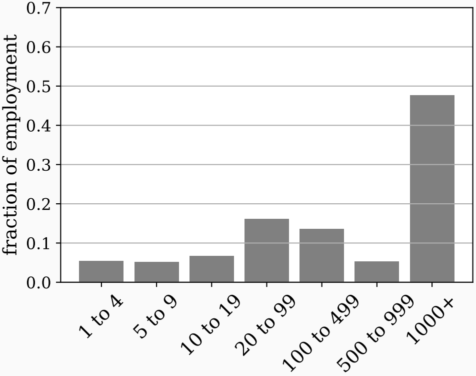
Distribution of firm size in 2022



- Highly dispersed distribution
- Majority are very small firms (1–4 employees)
- Over 10,000 firms with 1,000+ employees
- Over 1,000 firms with 10,000+ employees

Employment Shares

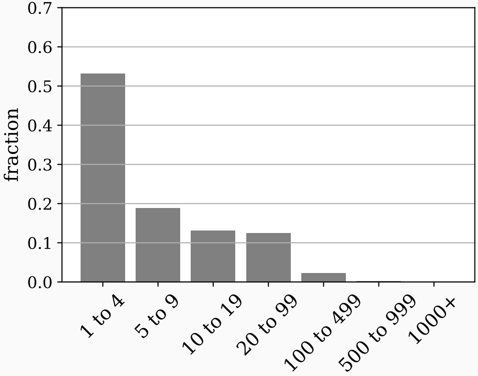
Employment share of each size category in 2022



- Small firms are numerous but large firms employ most workers
- $\approx 50\%$ of employees work at firms with $1{,}000+$ employees
- $\approx 30\%$ work at very large firms ($10{,}000+$ employees)

Establishment Size Distribution

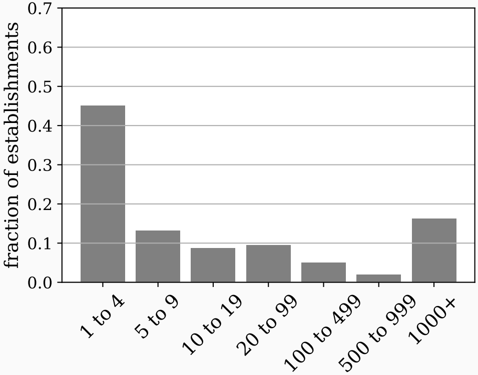
Distribution of establishment size in 2022



- $\approx 50\%$ are very small (1–4 employees)

Firm Size and Establishment Share

Fraction of establishments owned by each firm size category in 2022



- Large firms own many establishments
- Firms with $1{,}000+$ employees own avg ≈ 100 establishments
- Firms with $10{,}000+$ employees own avg ≈ 600 establishments

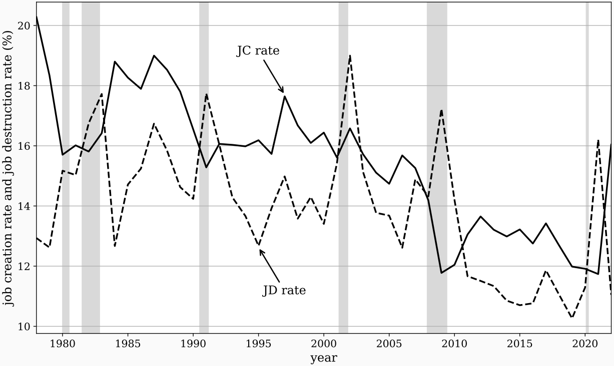
Job Creation and Destruction

Definitions: Let ℓ_{it} be employment of establishment i at year t

$$JC_t \equiv \frac{\sum_{i:\ell_{it}>\ell_{i,t-1}}(\ell_{it}-\ell_{i,t-1})}{\bar{L}_t}, \qquad JD_t \equiv \frac{\sum_{i:\ell_{it}<\ell_{i,t-1}}(\ell_{i,t-1}-\ell_{it})}{\bar{L}_t}$$

where $\bar{L}_t = (L_t + L_{t-1})/2$.

Annual job creation and destruction rates (establishment level)



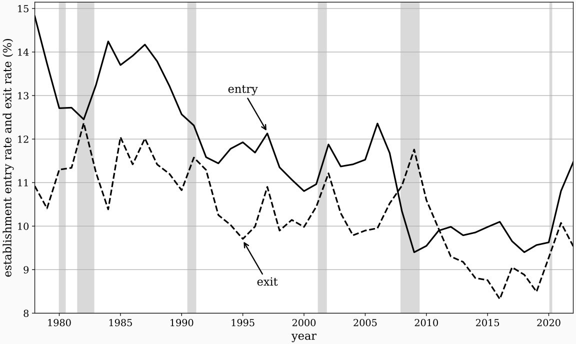
Properties of Job Flows

Three notable properties:

1. **Large magnitudes**: Both JC and JD rates exceed 10% in any given year
2. **Cyclical**: In recessions, JC declines and JD increases
3. **Declining trend**: Both rates have declined over time $\Rightarrow$ **"declining business dynamism"**

Entry and Exit Rates

Annual establishment entry and exit rates

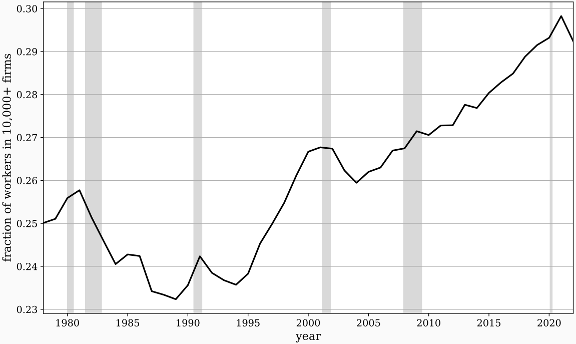


Similar properties as JC/JD:

- Large rates (entry ≈ 10–14%, exit ≈ 8–11%)
- Cyclical patterns
- Declining trend over recent decades

Rise of Large Firms

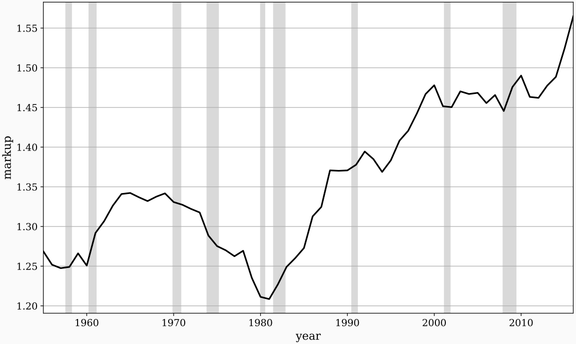
Fraction of employees working in firms with 10,000+ employees



- Steady increase since early 1990s
- Large firms are increasingly dominating the U.S. economy
- Raises concerns about **market power**

Rising Markups

Markup of U.S. public firms



- Average markup (price/marginal cost) of U.S. public firms
- Increasing trend since the 1980s
- Consistent with rising dominance of large firms
- Active research area; see Miller (2025) and Syverson (2025) for surveys

Reallocation and Misallocation

How Much Does Reallocation Matter?

Foster, Haltiwanger, and Krizan (2001) decomposition:

Define output-weighted average productivity: $\bar{A}_t \equiv \sum_i s_{it} a_{it}$

where s_{it} is the output share of establishment i.

$$\Delta \bar{A}_t = \underbrace{\sum_{i \in C} s_{i,t-1} \Delta a_{it}}_{\text{within}} + \underbrace{\sum_{i \in C} (a_{i,t-1} - \bar{A}_{t-1}) \Delta s_{it}}_{\text{between}} + \underbrace{\sum_{i \in C} \Delta a_{it} \Delta s_{it}}_{\text{cross}}$$

$$+ \underbrace{\sum_{i \in N} s_{it} (a_{it} - \bar{A}_{t-1})}_{\text{entry}} - \underbrace{\sum_{i \in X} s_{i,t-1} (a_{i,t-1} - \bar{A}_{t-1})}_{\text{exit}}$$

C: continuing, N: new entrants, X: exiting establishments.

Decomposition Results

Five sources of aggregate productivity growth:

1. **Within**: each establishment raises its own productivity
2. **Between**: high-productivity establishments gain market share
3. **Cross**: interaction of productivity growth and share changes
4. **Entry**: entrants are above average
5. **Exit**: exiting establishments are below average

Key finding (U.S. manufacturing, 1977–1987):

- Only 45% from the within effect
- The remaining 55% from reallocation (between, cross, entry, exit)

$\Rightarrow$ **Reallocation is crucial for aggregate productivity growth**

Misallocation: Setup

Restuccia and Rogerson (2008), Hsieh and Klenow (2009): firm-specific distortions cause misallocation.

Add firm-specific output tax τ_i to the simple model:

$$\max_{m_i} (1-\tau_i) a_i m_i^\gamma - c m_i$$

GDP is still $Y = \int y_i di$. Aggregate production function: $Y = AF(\mathbf{X})^\gamma$

But now:

$$A = \frac{\left[\int a_i^{\frac{1}{1-\gamma}} (1-\tau_i)^{\frac{\gamma}{1-\gamma}}\, di\right]^{1-\gamma}}{\left[\int a_i^{\frac{1}{1-\gamma}} (1-\tau_i)^{\frac{1}{1-\gamma}}\, di\right]^{\gamma}}$$

Reduces to the undistorted A when $\tau_i = 0$ for all i.

Misallocation: Lognormal Case

Assume $(a_i, 1-\tau_i)$ are bivariate lognormal:

$$(\log a_i, \log(1-\tau_i)) \sim N(\mu, \Sigma)$$

with $\mu = (\nu_a - \sigma_a^2/2,\ \nu_\tau - \sigma_\tau^2/2)$ and $\text{Var}(\log(1-\tau_i)) = \sigma_\tau^2$.

Result:

$$A = \exp\left(\nu_a + \frac{\gamma}{1-\gamma}\frac{1}{2}(\sigma_a^2 - \sigma_\tau^2)\right)$$

Key insights:

- Higher dispersion in taxes σ_τ **reduces** aggregate productivity
- Productive firms may not expand if $(1-\tau_i)a_i$ is low
- Unproductive firms may over-expand if $(1-\tau_i)a_i$ is high
- The level of the tax ν_τ does not affect A (no distortion in allocation)

Sources of Misallocation

Specific policies and institutions that cause misallocation:

- Size-specific taxes and regulations
- Entry regulations and barriers
- Regulations governing hiring and firing
- Financial frictions (credit constraints)
- Trade barriers

Quantitative importance: Hsieh and Klenow (2009) estimate that reallocating resources to equalize marginal products would raise manufacturing TFP by:

- 30–50% in China
- 40–60% in India

Firm Heterogeneity in General Equilibrium

The Hopenhayn-Rogerson (1993) Framework

Goal: Analyze reallocation frictions in general equilibrium with forward-looking firms.

Key features:

- Continuum of firms with idiosyncratic productivity shocks
- Endogenous entry and exit
- Labor as the only input
- Free entry condition
- Representative consumer
- Focus on steady state

Experiments: firing taxes and entry barriers (Moscoso Boedo and Mukoyama, 2012).

Production and Profits

- Production function: $y_t = a_t \ell_t^\gamma$
- a_t: idiosyncratic productivity (stochastic)
- Firms pay:
  - Wages: $w_t \ell_t$
  - Fixed operation cost: c_f (units of goods, per period)
  - Firing taxes: $\tau \max(0, \ell_{t-1} - \ell_t)$

Flow profit:

$$\pi(\ell_{t-1}, \ell_t, a_t) = a_t \ell_t^\gamma - w_t \ell_t - c_f - \tau \max(0, \ell_{t-1} - \ell_t)$$

- Output price normalized to 1
- Firing taxes transferred back to consumers

Timing and Productivity Process

Within-period timing:

1. Incumbent decides whether to **exit**
   - If exit: pay firing cost $\tau\ell_{t-1}$
2. If stay: receive current a_t from

$$\log(a_t) = \alpha + \rho \log(a_{t-1}) + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2)$$

3. Choose employment ℓ_t, produce

Dynamic element: With $\tau > 0$, hiring is dynamic—firm foresees future firing costs $\Rightarrow$ reluctance to hire even with positive shocks.

Firm's Dynamic Program

$$W(a, \ell_{-1}) = \max_{\ell} \; \pi(\ell_{-1}, \ell, a) + \beta \max \left\{ \mathbb{E}[W(a', \ell) \mid a], \; -\tau\ell \right\}$$

- $\beta \in (0, 1)$: consumer's discount factor
- Inner $\max$: **exit decision**
    - Continue: get $\mathbb{E}[W(a', \ell) \mid a]$
    - Exit: pay firing cost $-\tau\ell$

Entry and Consumer

Free entry:

$$W^e = c_e + \kappa$$

where c_e = technological entry cost, κ = policy-related entry barriers.

$$W^e = \int (W(a,0) + c_f)\, d\nu(a)$$

$\nu(a)$: exogenous distribution for entrant productivity.

Representative consumer:

$$\sum_{t=0}^{\infty} \beta^t [u(C_t) - \chi L_t^s]$$

Steady-state FOC: $wu'(wL^s + \Pi + R) = \chi$

where Π = total firm profits, R = total transfers.

Block-Recursive Structure

Key computational insight (Kaas, 2021):

1. Solve $W(a, \ell_{-1})$ as function of w ($\Rightarrow$ get $W^e(w)$)
2. Free entry pins down w^* ($\Rightarrow W^e(w^*) = c_e + \kappa$)
3. Given w^* and entry mass M: compute stationary distribution
4. Compute $L^d(M)$, $\Pi(M)$, $R(M)$
5. Labor market clearing $L^d(M) = L^s(\Pi(M), R(M))$ pins down M

Advantage: No need to track firm distribution to solve for prices ("block recursive structure")

$\Rightarrow$ Much easier computationally than heterogeneous-consumer models. (Lee and Mukoyama, 2018)

Calibration

- One period = 5 years
- $u(c) = \log(c)$
- $\beta = 0.8$ (4% annual discount rate)
- $\gamma = 0.64$ (labor share)
- Baseline: $(\tau, \kappa) = (0, 0)$ calibrated to the US economy

Targets:

- $\alpha = 0.076$: average firm size ≈ 62 employees
- $\rho = 0.93$: autocorrelation of $\log(\ell)$ is 0.93
- $\sigma = 0.253$: variance of employment growth is 0.53
- $c_f = 18.0$: exit rate $\approx 37\%$ (over 5 years)
- $c_e = 9.04$: free entry holds with $w = 1$

Effects of Firing Taxes

	$\tau = 0$	$\tau = 0.1$	$\tau = 0.2$
Wage	1.000	0.977	0.957
Total output	100	97.7	95.7
Total employment	100	98.3	97.4
Labor productivity	100	99.4	98.3
JC (= JD) rate	0.28	0.25	0.21

Table 1: Model results with firing taxes ($\kappa = 0$).

- $\tau = 0.1 \approx$ 6 months' salary
- $\tau = 0.2 \approx$ 1 year's salary

Interpreting the Firing Tax Results

1. **Employment declines**: Hiring effect dominates firing effect
   - Firms fire less $\Rightarrow$ pushes L up
   - Firms hire less (foresee future firing costs) $\Rightarrow$ pushes L down
   - Net: employment falls
2. **Labor productivity (Y/L) declines**: Misallocation
   - Good-a firms don't expand enough
   - Bad-a firms don't contract enough
   - Marginal products of labor become dispersed
3. **JC rate declines**: Less reallocation
   - Reluctance to hire and fire
   - Closely linked to productivity loss from misallocation

Effects of Entry Barriers

	$\kappa = 0$	$\kappa = 0.5$	$\kappa = 5.0$
Wage	1.000	0.986	0.879
Total output	100	98.6	87.9
Total employment	100	99.5	96.2
Labor productivity	100	99.1	91.4
JC (= JD) rate	0.28	0.28	0.28

Table 2: Model results with entry barriers ($\tau = 0$).

- $\kappa = 5.0$: comparable to low-income country entry costs
- Substantial output and productivity losses

Interpreting the Entry Barrier Results

Mechanism: Higher $\kappa \Rightarrow$ higher firm value in equilibrium $\Rightarrow$ lower wage

Three channels through which low wage affects productivity:

1. Low-productivity firms **less likely to exit** $\Rightarrow$ drags down A $\downarrow$
2. Low exit rate $\Rightarrow$ low entry rate $\Rightarrow$ fewer (less productive) entrants $\Rightarrow$ raises A $\uparrow$
3. Larger incumbent size $\Rightarrow$ DRS $\Rightarrow$ lower productivity $\downarrow$

Net effect: productivity **declines** (channels 1 & 3 dominate).

Political economy: Entry barriers **harm entrants** but **benefit incumbents** $\Rightarrow$ conflict of interest (Mukoyama and Popov, 2014).

Alternative Market Arrangements

Moving Beyond Perfect Competition

Two takeaways:

1. Misallocation insights carry over with minor modifications
2. Market power enables analysis of firm heterogeneity and **markups**

We consider two alternatives:

- **Monopolistic competition** (constant markups)
- **Oligopoly** (endogenous markups) — Atkeson and Burstein (2008)

Monopolistic Competition: Setup

Final good (CES aggregation / Dixit–Stiglitz):

$$Y = \left(\int y_i^{\frac{\sigma-1}{\sigma}} \, di \right)^{\frac{\sigma}{\sigma-1}}, \qquad \sigma > 1$$

- Each intermediate good i produced by a monopolist
- Same input structure as the simple model
- Input markets: perfectly competitive

Inverse demand for good i:

$$p_i = y_i^{-1/\sigma} Y^{1/\sigma}$$

Monopolistic Competition: Firm Problem

Producer i solves:

$$\max_{m_i} (a_i m_i^\gamma)^{-1/\sigma} Y^{1/\sigma} \cdot a_i m_i^\gamma - c m_i$$

taking Y as given (firm is small relative to aggregate).

Key result — Constant markup pricing:

$$p_i = \frac{\sigma}{\sigma - 1} \cdot \mathcal{M}$$

where $\mathcal{M}$ is marginal cost. The markup $\sigma/(\sigma-1)$ is **constant**.

Aggregate production function: $Y = AF(\mathbf{X})^\gamma$ with modified A:

$$A \equiv \left(\int a_i^{\frac{\sigma}{\sigma-1} - \gamma} \, di \right)^{\frac{\sigma-1}{\sigma-1-\gamma}}$$

Monopolistic Competition: Key Properties

- Aggregate production function same form as before
- Aggregation uses $\sigma/(\sigma-1)$ instead of $1/(1-\gamma)$
- **Advantage**: Can accommodate **constant** or even some **increasing** returns to scale
    - Monopoly power provides another force limiting firm size
- **Limitation**: Constant markup $\Rightarrow$ cannot analyze endogenous markup changes

Oligopoly: Atkeson-Burstein (2008)

Motivation: Need endogenous markups to study:

- Rising markups since the 1980s
- How changes in competition affect markups

Structure: Two layers of aggregation

1. **Across sectors**: CES with elasticity σ (same as before)

$$Y = \left(\int y_i^{\frac{\sigma-1}{\sigma}} \, di \right)^{\frac{\sigma}{\sigma-1}}$$

2. **Within sector** i: J firms ("brands") with elasticity η

$$y_i = \left(\sum_{j=1}^{J} q_{ij}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}$$

Assumption: $\eta > \sigma > 1$ (brands more substitutable than sectors).

Endogenous Markup Formula

Pricing rule (Cournot competition):

$$\hat{p}_{ij} = \frac{\varepsilon(s_{ij})}{\varepsilon(s_{ij}) - 1} \cdot \mathcal{M}$$

where s_{ij} is the **sales share** of firm j in sector i:

$$s_{ij} = \frac{\hat{p}_{ij} q_{ij}}{\sum_h \hat{p}_{ih} q_{ih}}$$

and the effective elasticity is:

$$\varepsilon(s_{ij}) = \left(\frac{1 - s_{ij}}{\eta} + \frac{s_{ij}}{\sigma} \right)^{-1}$$

- $\varepsilon = \eta$ when $s_{ij} = 0$ (small firm $\Rightarrow$ low markup)
- $\varepsilon = \sigma$ when $s_{ij} = 1$ (monopoly $\Rightarrow$ high markup)
- Monotonically decreasing in s_{ij}

Endogenous Markups: Properties

Symmetric firms: $s_{ij} = 1/J$

$$\varepsilon = \left(\frac{1}{\eta} \cdot \frac{J-1}{J} + \frac{1}{\sigma} \cdot \frac{1}{J}\right)^{-1}$$

- $J = 1$ (monopoly): $\varepsilon = \sigma \Rightarrow$ same as monopolistic competition
- $J \to \infty$: $\varepsilon \to \eta \Rightarrow$ most competitive case
- **Fewer firms** $\Rightarrow$ **higher markups**

Asymmetric firms: firm heterogeneity in a_{ij} feeds into markup heterogeneity through sales shares.

Bertrand competition: Same formula but with

$$\varepsilon(s_{ij}) = \eta(1 - s_{ij}) + \sigma s_{ij}$$

Business Cycles and Heterogeneous Firms

Do Idiosyncratic Shocks Wash Out?

Law of Large Numbers argument:

Let GDP $Y_t = \sum_{i=1}^{N} y_{it}$, with i.i.d. growth rates: $(y_{i,t+1} - y_{it})/y_{it} = \sigma \varepsilon_{i,t+1}$

Standard deviation of GDP growth:

$$\sigma_Y = \sigma \sqrt{\sum_{i=1}^{N} \left(\frac{y_{it}}{Y_t} \right)^2}$$

Equal-size firms: $y_{it}/Y_t = 1/N$

$$\sigma_Y = \frac{\sigma}{\sqrt{N}}$$

With $N = 5{,}000{,}000$ and $\sigma \approx 15\%$: $\sigma_Y \approx 0.00007 \Rightarrow$ **negligible**!

Aggregate Shocks and Firm Dynamics

Standard approach: Accept LLN, add aggregate shock z_t:

$$y_{it} = z_t \, s_{it} \, \ell_{it}^{\gamma}$$

where $z_t =$ aggregate productivity (as in RBC models).

Results:

- Modified Hopenhayn-Rogerson model replicates well:
  - Cyclical JC and JD rates
  - Cyclical entry and exit rates
- Block-recursive structure $\Rightarrow$ computationally tractable
- Some firm-level statistics harder to match (Lee and Mukoyama, 2018)

Idiosyncratic Productivity and Aggregate: Hulten's Theorem

Setup: N sectors, sector i produces $y_i = a_i F(k_i, \ell_i, x_{i1}, \ldots, x_{iN})$

where x_{ij} = intermediate input from sector j.

Theorem (Hulten, 1978): The first-order output effect of productivity changes:

$$\frac{dY}{Y} = \sum_i D_i \frac{da_i}{a_i}$$

where D_i is the **Domar weight**:

$$D_i = \frac{p_i y_i}{\sum_i p_i c_i} = \frac{\text{Sales of sector } i}{\text{GDP (value added)}}$$

Note: Domar weights use *sales* (not value added) ⇒ can exceed 1!

⇒ Downstream firms have amplified impact through input-output linkages.

Hulten's Theorem: Intuition

Two key intuitions:

1. **Why only a_i matters** (not inputs)?
   - Envelope theorem: in efficient economy, input adjustments have no first-order welfare effect
2. **Why sales (not value added)?**
   - With input-output networks, TFP improvement in a downstream firm also raises the value of intermediate inputs

Limitations (relaxed by Baqaee and Farhi, 2019, 2020):

- Requires efficient economy (no misallocation)
- First-order approximation only

Hulten's Theorem: Simple Example

Two sectors:

- Sector 1 (consumption): $y_1 = a_1 x^{1-\gamma} \ell^\gamma$
- Sector 2 (intermediate): $y_2 = a_2 k$

Fixed K, L. Competitive equilibrium: $x = y_2$

$$Y = a_1 (a_2 K)^{1-\gamma} L^\gamma$$
$$\frac{dY}{Y} = \underbrace{1}_{D_1} \cdot \frac{da_1}{a_1} + \underbrace{(1-\gamma)}_{D_2} \cdot \frac{da_2}{a_2}$$

- $D_1 = 1$ (sales = GDP for downstream sector)
- $D_2 = 1 - \gamma$ (value added share of upstream sector)
- Domar weights sum to $1 + (1-\gamma) > 1$

Large Firms and “Granular” Fluctuations

Gabaix (2011): Fat-tailed firm size distribution changes the LLN result.

If firm sizes follow a **Pareto distribution**: $\Pr[y_i > x] = \chi x^{-\zeta}$ (the US data: $\zeta \approx 1$, called the Zipf's law)

Result: Instead of $\sigma_Y \propto \sigma/\sqrt{N}$:

$$\sigma_Y \sim \frac{v_\zeta \, \sigma}{\log(N)}$$

- $1/\log(1{,}000{,}000) \approx 0.072$ vs. $1/\sqrt{1{,}000{,}000} = 0.001$
- **Two orders of magnitude larger!**
- Idiosyncratic shocks to large firms can generate aggregate-scale fluctuations

Carvalho & Grassi (2019): quantitative business cycle model driven by idiosyncratic shocks to large firms.

Production Networks

Additional amplification: Domar weights $D_i = p_i y_i / \text{GDP}$

- Large downstream firms (Walmart, Amazon, GM): sales $\gg$ value added
- Their Domar weights can be much larger than their GDP share
- $\Rightarrow$ Idiosyncratic shocks amplified through input-output linkages

Beyond Hulten's theorem:

- **Baqaee & Farhi (2019)**: Second-order effects can be quantitatively important
- **Baqaee & Farhi (2020)**: With distortions, network structure matters for first-order effects; decompose Solow residual into misallocation changes + "pure" technology growth

Endogenous Productivity:
Klette-Kortum (2004)

Motivation

- Previous sections: productivity a_i is **exogenous**
- Natural question: What determines a_i?
- **Klette & Kortum (2004)**: Endogenous productivity from **innovation**
- Combines:
    - Quality-ladder growth models (Ch. 13)
    - Firm dynamics (entry, exit, expansion, contraction)
- Major strength: clear definition of a "firm" and its dynamics

Preferences

Representative consumer:

$$\sum_{t=0}^{\infty} \beta^t \log(C_t)$$

Consumption aggregator (unit elasticity of substitution):

$$C_t = \exp\left(\int_0^1 \log\left(\sum_{k=-1}^{J_t(j)} q_t(j,k)\, c_t(j,k)\right) dj\right)$$

- $j \in [0,1]$: product index
- k: generation (quality rung), $J_t(j) =$ cutting-edge generation
- $q_t(j,k)$: quality of product (j,k)
- $c_t(j,k)$: consumption of product (j,k)
- Different generations are **perfect substitutes** (quality-adjusted)

Intratemporal Problem

Equal expenditure shares (unit elasticity):

$$c_t(j,k) = \frac{E_t}{p_t(j,k)}$$

for each j, consumer buys only the generation with lowest quality-adjusted price.

Price index: $P_t = \exp\left(\int_0^1 [\log p_t(j,k) - \log q_t(j,k)]\, dj\right)$

Normalize $P_t = 1$.

Intertemporal Euler equation (balanced growth path):

$$\frac{1}{C_t} = \beta(1+r)\frac{1}{C_{t+1}} \quad \Rightarrow \quad \frac{1}{1+r} = \frac{\beta}{1+\gamma}$$

where γ is the consumption growth rate.

Production and Pricing

Production: 1 unit of labor $\Rightarrow$ 1 unit of any product.

- Unit cost $= w_t$ for all products

Limit pricing: The cutting-edge producer sets:

$$p_t(j, J_t(j)) = \lambda w_t$$

where $\lambda > 1$ is the **quality step** (also the markup).

Period profit per product line:

$$\pi_t = \left(1 - \frac{1}{\lambda}\right) C_t$$

- All product lines earn the same profit.
- Firm size $=$ number of product lines owned

Innovation

Innovation intensity: η per product line

- Cost: $w_t R(\eta)$ units of labor, $R(\cdot)$ increasing and convex
- Success probability: η
- If successful: gain a random product line (quality λ times the current best of that line)
- Fixed total number of products $\rightarrow$ New product always taken from another firm (creative destruction)

Let $\mu =$ probability of **losing** a product line to another innovator.

Bellman equation (per product line):

$$V_t = \max_{\eta} \; \pi_t - w_t R(\eta) + \frac{1}{1+r}(1+\eta-\mu)V_{t+1}$$

Expected continuation: $(1+\eta-\mu)V_{t+1}$ (gain with prob η, lose with prob μ).

Balanced Growth Path

Divide both sides by $(1+\gamma)^t$, use $\beta/(1+\gamma) = 1/(1+r)$ and call $v \equiv V_t/(1+\gamma)^t$:

$$v = \max_{\eta} \left(1 - \frac{1}{\lambda}\right) C_0 - R(\eta) + \beta(1+\eta-\mu)v,$$

FOC: $R'(\eta) = \beta v$

Free entry: An entrant pays c_e units of labor, gets one product line:

$$v = c_e$$

Creative destruction: $\mu = \eta + \nu$ (incumbent innovation + entry rate)

Labor market clearing:

$$\frac{C_0}{\lambda} + R(\eta) + \nu c_e = L$$

Equilibrium

Four unknowns: v, η, C_0, μ

Four equations:

$$v = \left(1 - \frac{1}{\lambda}\right) C_0 - R(\eta) + \beta(1 + \eta - \mu)v$$

$$R'(\eta) = \beta v$$

$$v = c_e$$

$$\frac{C_0}{\lambda} + R(\eta) + \nu c_e = L \qquad (\nu = \mu - \eta)$$

Growth rate: $\gamma = \mu \log(\lambda)$

- Growth driven by creative destruction rate μ and innovation step λ
- Both incumbent innovation η and entry ν contribute to growth

Firm Dynamics and Gibrat's Law

Firm size $=$ number of product lines

Average growth rate of the firm size (number of product lines per firm): $\eta - \mu = -\nu < 0$

- Net growth rate is **negative** (because entry displaces incumbents)
- Same for all firms regardless of size $\Rightarrow$ **Gibrat's Law**

Advantage over exogenous-shock models:

- Policies affect the productivity process itself
- Can analyze how innovation incentives shape firm dynamics
- Clear firm definition through product lines

Generating a Pareto Firm-Size Distribution

Problem: Basic Klette–Kortum doesn't produce a Pareto tail (average growth is negative).

Fix: Suppose large firms have positive constant growth rate $g > 0$, with exit probability δ. When the density at size n is $h(n)$,

Stationarity requires: $(1 + g)h((1 + g)n) = (1 - \delta)h(n)$

Guess Pareto distribution: $h(n) = Fn^{-(\zeta+1)} \Rightarrow$

$$\zeta = -\frac{\log(1 - \delta)}{\log(1 + g)}$$

- ζ small (thick tail) when δ small or g large
- **Modification of the Klette-Kortum model to have** $g > 0$: Allow some innovation to create **new** products.
- If new product creation ξ is large enough:
  $g = \eta - \mu = \xi - \mu > 0$

Counterfactual Predictions and Extensions

Limitations of the basic model:

1. All product lines earn equal profits
   - Unit elasticity $\Rightarrow$ revenue independent of quality
   - Fix: higher substitutability ($\sigma > 1$)
2. No cumulative innovation advantage for incumbents
   - Any firm can innovate at the same cost
   - Fix: allow incumbents to improve own products
3. No Pareto tail in firm-size distribution
   - Fix: allow new product creation alongside creative destruction
   - Luttmer (2011) discusses related insights

Summary

Key Takeaways

1. **Aggregation matters**: Firm heterogeneity affects aggregate productivity through endogenous resource allocation
2. **Data**: U.S. shows dispersed firm sizes, large reallocation, declining dynamism, rising concentration and markups
3. **Misallocation**: Firm-specific distortions reduce aggregate TFP; dispersion in distortions (not level) is what matters
4. **General equilibrium**: Firing taxes and entry barriers have substantial effects on output, productivity, and reallocation
5. **Market power**: Monopolistic competition gives constant markups; oligopoly (Atkeson-Burstein) generates endogenous markups linked to market shares
6. **Business cycles**: Fat tails and production networks can amplify firm-level shocks to aggregate fluctuations
7. **Endogenous productivity**: Klette-Kortum links innovation, creative destruction, and firm dynamics