Heterogeneous Firms in Macroeconomics

Toshihiko Mukoyama
Georgetown University

July 2022 (revised April 2024)

The textbook project

- Based on a chapter of the Ph.D.-level textbook I am team-writing with Marina Azzimonti, Per Krusell, Alisdair McKay, and others.

The plan

1. Introduction: why heterogeneity matters
2. Firm heterogeneity in the U.S. data
3. Reallocation and misallocation
4. Firm heterogeneity in general equilibrium
5. Alternative market arrangements
6. Business cycles and heterogeneous firms
7. Endogenous productivity

Introduction

A simple model

- Production function for firm i:

$$y_i = a_i F(\mathbf{x}_i)^\gamma$$

  The productivity a_i can be heterogeneous.
  $F(\cdot)$ is constant returns and $\gamma \in (0, 1)$: decreasing returns to scale. How does the a_i heterogeneity matter in the aggregate?
- Optimization in two steps: first, cost minimization (common for all firms)

$$\min_{\mathbf{x}} \ \mathbf{px}$$

  subject to

$$F(\mathbf{x}) = 1,$$

  with solution $\mathbf{x}^*$ and $c = \mathbf{px}^*$.

Introduction

- The second step: Let $m_i = F(\mathbf{x}_i^*)$ be the choice of the firm i's "combined inputs."
- The profit maximization problem:

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

From the first-order condition

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

$y_i = (c/\gamma) m_i$ holds.
- The production function aggregates to:

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

where

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

Thus the distribution of a_i influences A.

Introduction

- An example: a_i follows a lognormal distribution

$$\ln(a_i) \sim N(\nu - \sigma^2/2, \sigma^2).$$

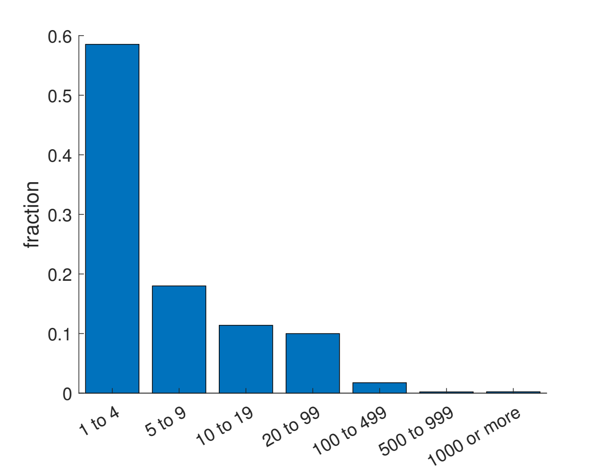
Then, the aggregate productivity A is

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

The increase in σ does not influence the mean of a_i in its distribution, but increases A. The effect of σ is larger when γ is closer to one, because highly productive firms can scale larger.

U.S. facts

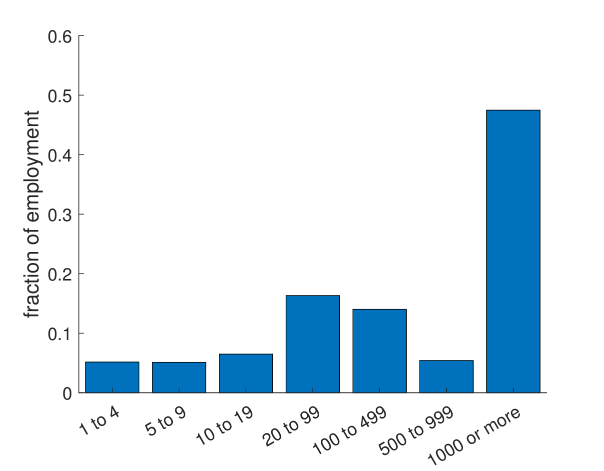
Distribution: Firm size measured by employment



Source: BDS (U.S. Census Bureau)

U.S. facts

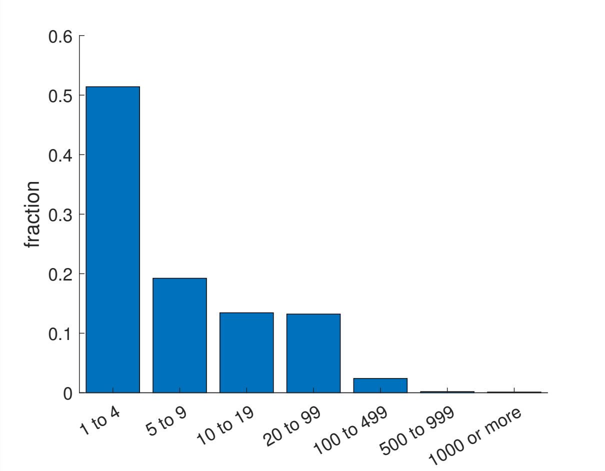
Fraction of people employed by each category



Source: BDS (U.S. Census Bureau)

U.S. facts

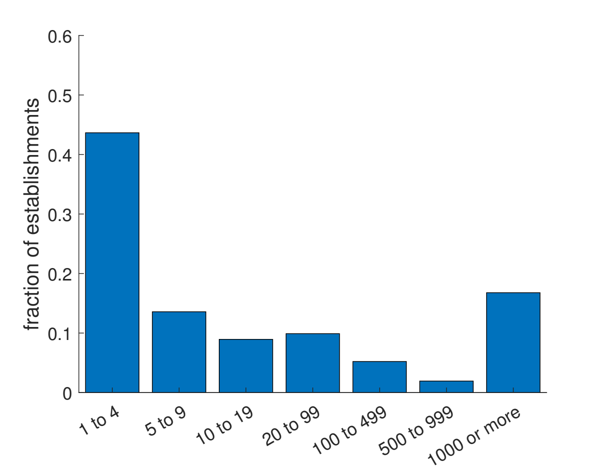
Establishment size measured by employment



Source: BDS (U.S. Census Bureau)

U.S. facts

Number of establishments at each firm



Source: BDS (U.S. Census Bureau)

U.S. facts

Reallocation: Job creation and job destruction rates

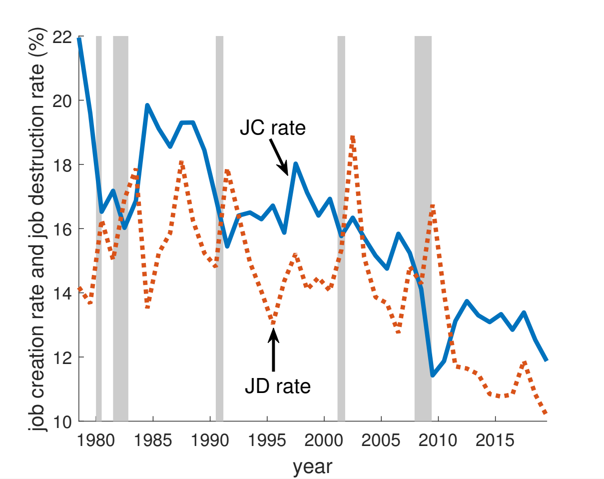
$$JC_t \equiv \frac{\sum_{i:\ell_{it}>\ell_{i,t-1}}(\ell_{it} - \ell_{i,t-1})}{\bar{L}_t},$$

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

These statistics measure the (gross) expansion and contraction of establishments (or firms).

U.S. facts

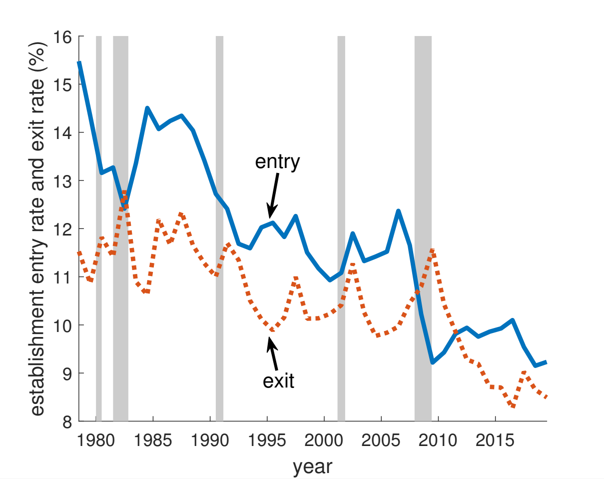
Job creation rate and job destruction rate (establishments)



Source: BDS (U.S. Census Bureau)

U.S. facts

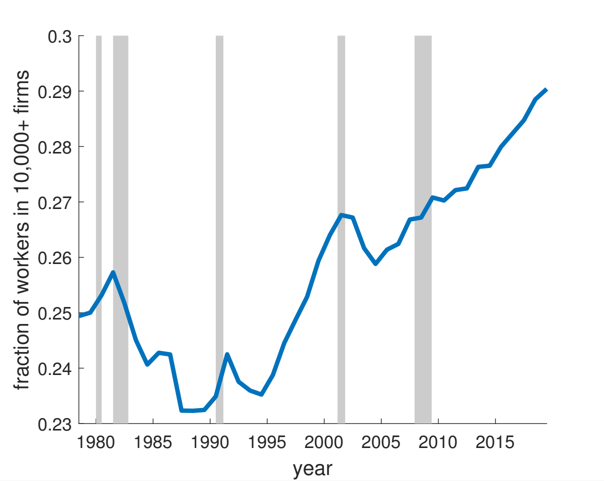
Entry rate and exit rate (establishments)



Source: BDS (U.S. Census Bureau)

U.S. facts

The fraction of employees working at the 10,000+ employee firms



Source: BDS (U.S. Census Bureau)

Reallocation and misallocation

- Foster, Haltiwanger, and Krizan (2001) decomposition:

$$\bar{A}_t \equiv s_{it} a_{it},$$

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

$$\begin{aligned}\Delta \bar{A}_t = & \sum_{i \in C} s_{it-1} \Delta a_{it} + \sum_{i \in C} (a_{it-1} - \bar{A}_{t-1}) \Delta s_{it} + \sum_{i \in C} \Delta a_{it} \Delta s_{it} \\ & + \sum_{i \in N} s_{it} (a_{it} - \bar{A}_{t-1}) - \sum_{i \in X} s_{it-1} (a_{it-1} - \bar{A}_{t-1})\end{aligned}$$

- All factors other than the first factor is due to reallocation.
- Using the U.S. Manufacturing data from 1977 to 1987, Foster, Haltiwanger, Krizan (2001) estimate that the aggregate change in multifactor productivity is 45% accounted for by the first factor, and the rest of 55% is the contribution of reallocation.

Reallocation and misallocation

- "Misallocation" with idiosyncratic distortions
- Firm i is taxed at the idiosyncratic rate τ_i. The problem is now

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

The aggregate production function is still $Y = AF(\mathbf{X})^{\gamma}$ with

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

When $(\ln(a_i), \ln(1-\tau_i)) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\nu_a - \sigma_a^2/2, \nu_\tau - \sigma_\tau^2/2)$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_\tau \\ \rho\sigma_a\sigma_\tau & \sigma_\tau^2 \end{bmatrix}$

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

Firm heterogeneity in general equilibrium

Hopenhayn and Rogerson (1993): dynamic + general equilibrium

- The firm's flow profit (facing a firing tax τ)

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

- The idiosyncratic productivity changes over time:

$$\ln(a_t) = \alpha + \rho \ln(a_{t-1}) + \varepsilon_t,$$

$\varepsilon_t \sim N(0, \sigma^2)$.

- The firm's optimization

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

- Free entry:

$$W^e = c_e,$$

where

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

Firm heterogeneity in general equilibrium

- The representative consumer's problem in the steady state

$$\max_{C,L^s} \ u(C) - \chi L^s$$

subject to

$$C \leq wL^s + \Pi + R.$$

- The competitive equilibrium is "block recursive":
  - The wage w is determined by the firm's optimization and the free entry condition.
  - For a given entry mass, the stationary distribution of incumbents can be computed. The entry mass is determined so that $L^s = L^d$.
- Employment outcome: it is not a priori clear whether L increases with τ. (Firing $\downarrow$, but hiring also $\downarrow$ with forward-looking firms)
- Misallocation: Y/L declines with τ.

Alternative market arrangement: monopolistic competition

- The final good is produced by (Dixit-Stiglitz, CES)

$$Y = \left[\int y_i^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}} .$$

The cost minimization problem of a (competitive) final good producer

$$\min_{\{y_i\}} \int p_i y_i di$$

subject to the production function for a given Y.

$$p_i = \lambda y_i^{-\frac{1}{\sigma}} Y^{\frac{1}{\sigma}},$$

λ is the Lagrange multiplier for the production constraint, and it turns out it can be interpreted as the price of the final good. Normalize it to one.

Alternative market arrangement: monopolistic competition

- The intermediate-good producers are monopolists and solve

$$\max_{m_i} \; (a_i m_i^\gamma)^{-\frac{1}{\sigma}} Y^{\frac{1}{\sigma}} a_i m_i^\gamma - c m_i.$$

Each firm takes Y as given. In the Nash equilibrium among the monopolists, the same aggregation as before ($Y = AF(\mathbf{X})^\gamma$) holds, where

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

Because $\sigma/(\sigma - 1) > 1$, γ does not have to be less than one.

Alternative market arrangement: oligopoly and markups

- In the monopolistic competition case above, the markup turns out to be constant:
$$p_i = \frac{\sigma}{\sigma - 1}\mathcal{M},$$
where $\mathcal{M} \equiv \partial(cm_i)/\partial y_i$.
- Thus this framework cannot be used for analyzing the change in markups. There are many alternative formulations with variable markups, but here I will introduce the Cournot formulation based on Atkeson and Burstein (2008).
- Now there are two levels of nesting ("brands" within a "sector")
$$Y = \left[\int y_i^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}} \text{ and } y_i = \left[\sum_{j=1}^{J} q_{ij}^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}},$$
where $\eta > \sigma > 1$

Alternative market arrangement: oligopoly and markups

- Within a sector, a firm is "large" in the sense it is aware that q_{ij} can influence y_i. The optimization problem is

$$\max_{q_{ij},m_{ij}} \; q_{ij}^{-\frac{1}{\eta}} y_i^{\frac{1}{\eta}} y_i^{-\frac{1}{\sigma}} Y^{\frac{1}{\sigma}} q_{ij} - c m_{ij}$$

where

$$y_i = \left[\sum_{j=1}^{J} q_{ij}^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}}.$$

The solution is

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

where

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

Thus the markup is increasing in $s_{ij} \equiv \frac{\hat{p}_{ij}q_{ij}}{p_i y_i} = \frac{\hat{p}_{ij}q_{ij}}{\sum_{h=1}^{J}\hat{p}_{ih}q_{ih}}$.

Business cycles and heterogeneous firms

- With many firms, idiosyncratic shocks cancel out with each other (LLN).

$$\frac{y_{i,t+1} - y_{it}}{y_{it}} = \sigma \varepsilon_{i,t+1},$$

Then

$$\frac{Y_{t+1} - Y_t}{Y_t} = \frac{1}{Y_t} \sum_{i=1}^{N} \Delta y_{i,t+1} = \sum_{i=1}^{N} \frac{y_{it}}{Y_t} \sigma \varepsilon_{i,t+1}.$$

Thus the standard deviation of GDP growth rate is

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

which is $\sigma/\sqrt{N}$ if all firms are the same. With 1 million firms, $1/\sqrt{N} = 0.1\%$.

- One reaction: need an agg shock for business cycle analysis.
- Another reaction: maybe not all firms are the same.

Business cycles and heterogeneous firms

- Hulten's Theorem:

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

where D_i is the Domar weight (the numerator is sales):

$$D_i = \frac{p_i y_i}{\sum_i p_i c_i}.$$

- Gabaix (2011): when

$$\Pr[y_i > x] = \chi x^{-\zeta}$$

and $\zeta = 1$, then

$$\sigma_Y \sim \frac{v_\zeta}{\ln(N)} \sigma.$$

With 1 million firms, the coefficient is 7.2% instead of 0.1%. ("Granular dynamics")

- Production networks (sales $>>$ value added)

Endogenous productivity

Klette and Kortum (2004)

- Endogenous productivity (quality ladders) with firm dynamics.
- Consumers:

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

  where

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

- Intratemporal problem:
  - Purchase only generation with lowest "quality-adjusted price" $p_t(j,k)/q_t(j,k)$.
  - minimize expenditure $\rightarrow$

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

Endogenous productivity

- Thus
$$C_t = E_t \exp\left(\int_0^1 [\ln(q_t(j,k)) - \ln(p_t(j,k))]dj\right).$$
This relationship can be rewritten as $P_t C_t = E_t$, with the price index
$$P_t \equiv \exp\left(\int_0^1 [\ln(p_t(j,k)) - \ln(q_t(j,k))]dj\right).$$
Normalize $P_t = 1$.
- Intertemporal problem:
$$\max_{C_t} \sum_{t=0}^{\infty} \beta^t \ln(C_t)$$
subject to
$$\sum_{t=0}^{\infty} \left(\frac{1}{1+r}\right)^t C_t \leq \mathcal{A}_0,$$

Endogenous productivity

- Thus
$$C_t = E_t \exp\left(\int_0^1 [\ln(q_t(j,k)) - \ln(p_t(j,k))]dj\right).$$
This relationship can be rewritten as $P_t C_t = E_t$, with the price index
$$P_t \equiv \exp\left(\int_0^1 [\ln(p_t(j,k)) - \ln(q_t(j,k))]dj\right).$$
Normalize $P_t = 1$.
- Intertemporal problem:
$$\max_{C_t} \sum_{t=0}^{\infty} \beta^t \ln(C_t)$$
subject to
$$\sum_{t=0}^{\infty} \left(\frac{1}{1+r}\right)^t C_t \leq \mathcal{A}_0,$$

Endogenous productivity

- A firm produces and earns monopoly profit.

$$\pi_t \equiv (p_t(j, J_t(j)) - w_t)\frac{C_t}{p_t(j, J_t(j))} = \left(1 - \frac{1}{\lambda}\right) C_t.$$

- It innovates with the cost $w_t R(\eta)$, where η is the innovation intensity. It takes over another firm's product line when successfully innovates.
- Firm's optimization

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

can be normalized to

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

Note the unknowns: C_0, v, η, μ.

Endogenous productivity

The general equilibrium of the model:

- Entry: free entry

$$v = c_e.$$

- The total innovation is the sum of the incumbents' innovation and the entrants' innovation

$$\mu = \eta + \nu.$$

- The labor market equilibrium condition:

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

- The aggregate growth rate is $\mu \ln(\lambda)$.

Endogenous productivity

Firm dynamics:

- The expected value of the growth rate of a firm is $-\nu$.
  (Grows at the rate η, contracts with the rate $\mu = \eta + \nu$.)
- The model cannot generate a Pareto tail.

Endogenous productivity

An alternative setting that can generate a Pareto tail:

- A (large) firm has a positive constant growth rate g. All firms receives a exit shock with the probability $\delta \in (0,1)$.
- In the stationary distribution

$$(1+g)h((1+g)n)\Delta = (1-\delta)h(n)\Delta$$

  has to hold.
- Guess that the distribution is Pareto: $h(n) = Fn^{-(\zeta+1)}$. Then

$$(1+g)F((1+g)n)^{-(\zeta+1)}\Delta = (1-\delta)Fn^{-(\zeta+1)}\Delta.$$

  This equality holds for any n and Δ when

$$\zeta = -\frac{\ln(1-\delta)}{\ln(1+g)} > 0.$$

Endogenous productivity

How can we make the firm's average growth rate to be positive?

- For example, suppose that the *new product creation* among the total innovation is ξ (that is, among the total $\eta + \nu$ innovations, ξ create new products, and $\mu = \eta + \nu - \xi$ replace existing products).
- Then, the average growth rate of a firm, which is still $\eta - \mu$, is now equal to $\xi - \nu$ (instead of just $-\nu$). If ξ is sufficiently large, $\xi - \nu$ can be positive.