Endogenous Firm Structure and Worker Specialization

Jacob Adenbaum

University of Edinburgh

June 2023

The Division of Labor and the Extent of the Market

- Worker productivity within a firm depends on
  1. The skills of the workers hired
  2. How those workers are organized in production
- Example: Small Corner Shop vs. Big Supermarket
  - Same tasks (Sweep floors, stock shelves, staff register, manage inventories)
  - Different sets of workers (managers vs. managers, clerks, and janitors)
- Worker specialization depends on the firm's scale of production

Question: What is the value of worker specialization within the firm?

The Division of Labor and the Extent of the Market

- Worker productivity within a firm depends on
  1. The skills of the workers hired
  2. How those workers are organized in production
- Example: Small Corner Shop vs. Big Supermarket
  - Same tasks (Sweep floors, stock shelves, staff register, manage inventories)
  - Different sets of workers (managers vs. managers, clerks, and janitors)
- Worker specialization depends on the firm's scale of production

Question: What is the value of worker specialization within the firm?

The Division of Labor and the Extent of the Market

- Worker productivity within a firm depends on
  1. The skills of the workers hired
  2. How those workers are organized in production
- Example: Small Corner Shop vs. Big Supermarket
  - Same tasks (Sweep floors, stock shelves, staff register, manage inventories)
  - Different sets of workers (managers vs. managers, clerks, and janitors)
- Worker specialization depends on the firm's scale of production

Question: What is the value of worker specialization within the firm?

Approach

Data:

- Use Brazilian matched employer-employee data merged with occupation-specific skill measures from O*NET
- Novel Facts: Firms operating at different scales hire different types of workers
    1. Average skills vary non-monotonically in firm size
    2. As firms grow, they add more specialized occupations (more extreme distribution of skills)

Theory:

- New theory of how firms choose:
    1. Which occupations to hire (number of occupations, and which skills)
    2. How to assign tasks to workers (time allocation)
- Model generates endogenous **hierarchy of specialization** across firms

Approach

Data:

- Use Brazilian matched employer-employee data merged with occupation-specific skill measures from O*NET
- Novel Facts: Firms operating at different scales hire different types of workers
  1. Average skills vary non-monotonically in firm size
  2. As firms grow, they add more specialized occupations (more extreme distribution of skills)

Theory:

- New theory of how firms choose:
  1. Which occupations to hire (number of occupations, and which skills)
  2. How to assign tasks to workers (time allocation)
- Model generates endogenous **hierarchy of specialization** across firms

Approach

Estimation and Results:

- Novel Identification Strategy: use cross-sectional variation in firms' occupation-specific wage bill shares to identify the primitives of the task-based production function
- Estimate model using firms' FOCs for Brazil's manufacturing sector
- 36% of the variation in firm productivity is due to endogenous specialization channel
  - firms with a higher exogenous productivity hire a set of workers who are more productive at the tasks they are assigned
- Counterfactuals:
  - Gains from reducing cost of specialization are modest (1.2% increase in output).
  - Shutting down specialization channel results in 9.6% decline in manufacturing output

Related Literature

- **Task Assignment:** Rosen (1978), Acemoglu and Autor (2011), Ocampo (2019), Ales, Combemale, Fuchs, and Whitefoot (2021), Acemoglu and Restrepo (2021)

  Contribution: Identification Strategy based on firm level heterogeneity and occupation based skill measures

- **Firm Structure and Worker Specialization:** Rosen (1982), Garicano (2000), Caliendo and Rossi-Hansberg (2012), Caliendo, Monte, and Rossi-Hansberg (2015),

  Contribution: Allow for multidimensional skills. More flexible measurement strategy for worker characteristics

- **Returns to Skill and Technology:** Katz and Murphy (1992), Heckman and Sedlacek (1985), Autor and Dorn (2013), Buera, Kaboski, and Rogerson (2015), Lindenlaub (2017), Lise and Postel-Vinay (2020),

  Contribution: Novel theory of the firm's production function with heterogeneous labor; optimal mix of occupations depends on the scale of firm production.

Section 2

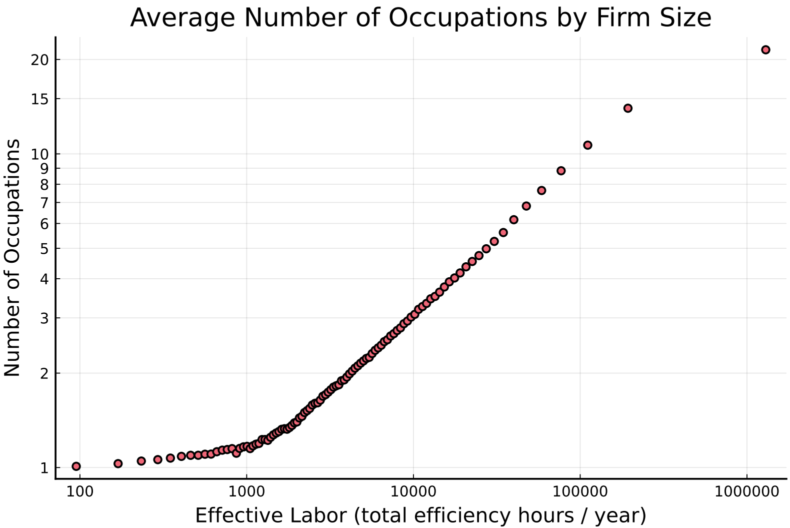
Data

Data Sources

1. Brazilian Administrative Data (RAIS)
   - Covers 1994 - 2010: all workers in Brazil's formal sector
   - Matched employer employee data: hours, industry, and occupation
2. O*NET Skill Data
   - Comprehensive database of occupation specific skill measures
   - Follow Lise and Postel-Vinay (2020) to calculate measures of Cognitive, Manual, and Interpersonal Skill
   - Skill measures vary between 0 and 1
   - Merge with Brazilian occupation codes using mapping from De Souza (2020)

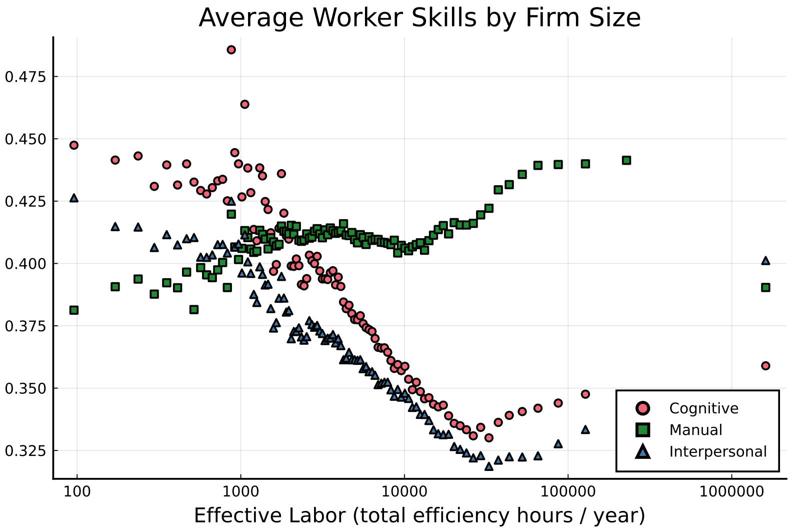
I document a set of novel facts about how **firms vary the composition of their workforce** with their size

Larger Firms Hire More Types of Workers



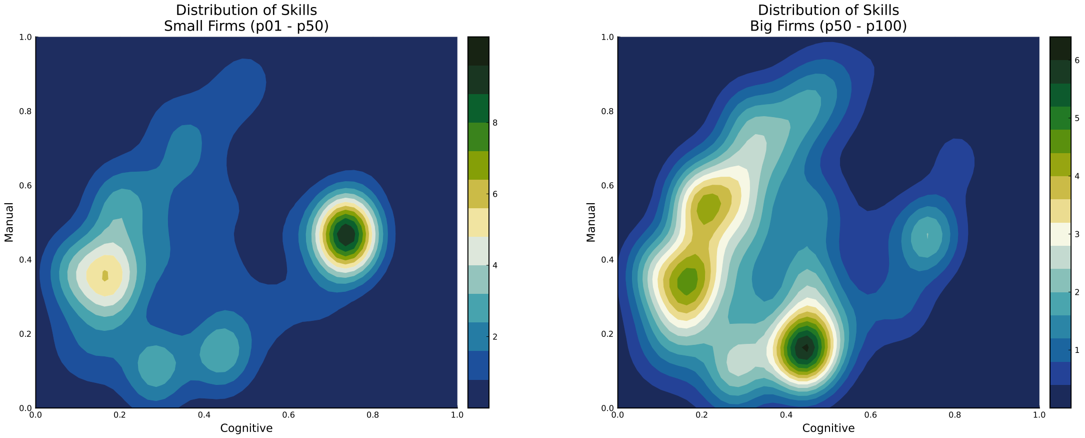
Total Labor | Wage Bill | Num Employees | Effective Labor Def

Larger Firms Hire Workers with Different Skills



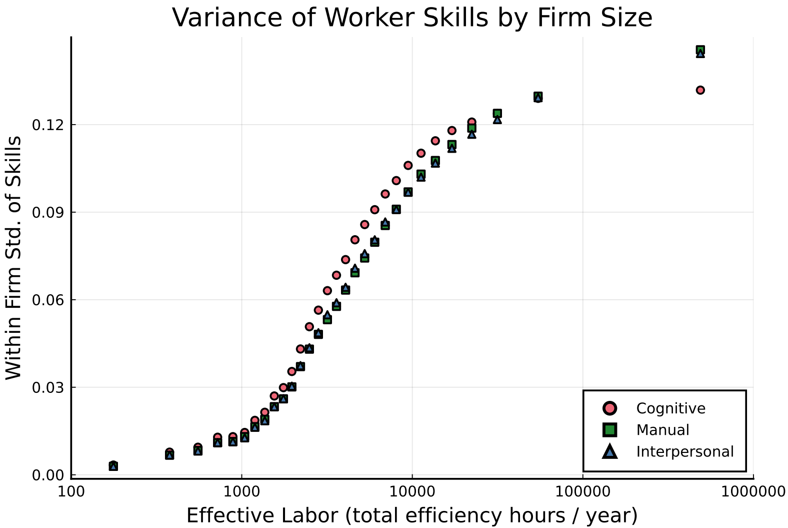
Total Labor | Composition Adjusted

Managers and Workers



C × I   M × I

Larger Firms Spread out their Workers More



Total Labor Wage Bill

Controlling for Industry Composition

- Interested in how each of the following varies with firm size
  - log(Occupations)
  - Average skills of workers (cognitive, manual, and interpersonal)
  - Within-firm standard deviation of skills (cognitive, manual, and interpersonal)
- Consider the regression:

$$y_i = \sum_{s=2}^{10} \beta_s D_i^s + \gamma_{d(i)} + \epsilon_i$$

  - $\gamma_{d(i)}$ is a fixed effect for each 5-digit industry code $d(i)$
  - D_i^s is an indicator for whether firm i is in decile s of the firm size distribution

Controlling for Industry Composition

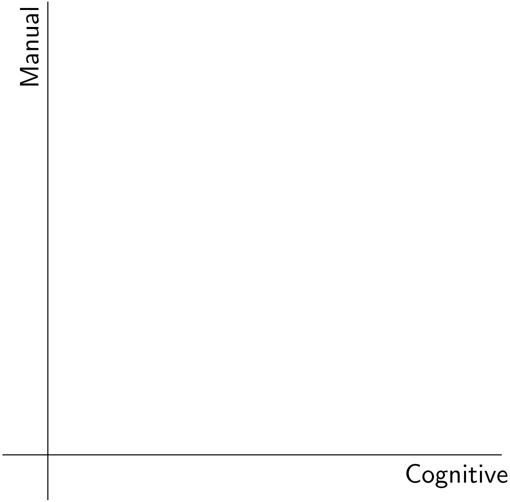
	Within Firm Std Skills				Avg Skills		
	log(Occupations)	Cognitive	Manual	Interpersonal	Cognitive	Manual	Interpersonal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
deciles: 2	0.051***	0.006***	0.006***	0.006***	0.003	0.011***	-0.004***
	(0.004)	(4.964e-04)	(4.360e-04)	(4.567e-04)	(0.002)	(0.001)	(0.002)
deciles: 3	0.119***	0.016***	0.014***	0.014***	-0.014***	0.014***	-0.019***
	(0.006)	(7.749e-04)	(9.592e-04)	(8.184e-04)	(0.003)	(0.002)	(0.002)
deciles: 4	0.220***	0.030***	0.026***	0.026***	-0.021***	0.018***	-0.025***
	(0.011)	(0.001)	(0.002)	(0.001)	(0.004)	(0.003)	(0.002)
deciles: 5	0.368***	0.051***	0.042***	0.043***	-0.025***	0.018***	-0.027***
	(0.016)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
deciles: 6	0.521***	0.069***	0.057***	0.058***	-0.030***	0.017***	-0.030***
	(0.019)	(0.003)	(0.003)	(0.002)	(0.003)	(0.005)	(0.003)
deciles: 7	0.706***	0.086***	0.074***	0.074***	-0.035***	0.015**	-0.032***
	(0.020)	(0.003)	(0.003)	(0.002)	(0.004)	(0.007)	(0.005)
deciles: 8	0.927***	0.102***	0.090***	0.089***	-0.038***	0.012	-0.033***
	(0.020)	(0.003)	(0.004)	(0.003)	(0.006)	(0.009)	(0.008)
deciles: 9	1.212***	0.116***	0.106***	0.104***	-0.043***	0.014*	-0.037***
	(0.021)	(0.003)	(0.004)	(0.003)	(0.006)	(0.008)	(0.007)
deciles: 10	1.846***	0.127***	0.123***	0.119***	-0.044***	0.022***	-0.038***
	(0.047)	(0.003)	(0.003)	(0.003)	(0.005)	(0.007)	(0.006)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1330135	1330135	1330135	1330135	1330135	1330135	1330135
R^2	0.632	0.326	0.369	0.388	0.344	0.410	0.352

Narrow Industries

Section 3

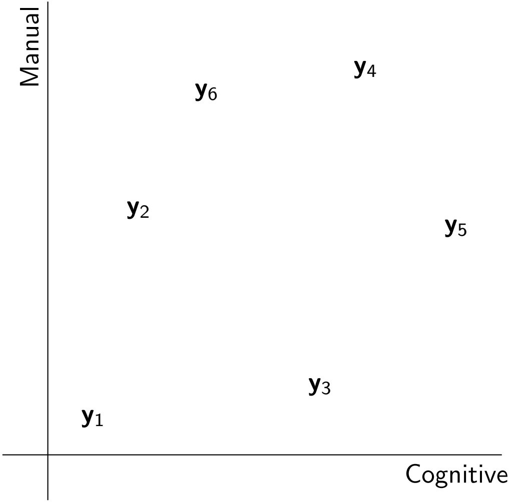
Model

Tasks



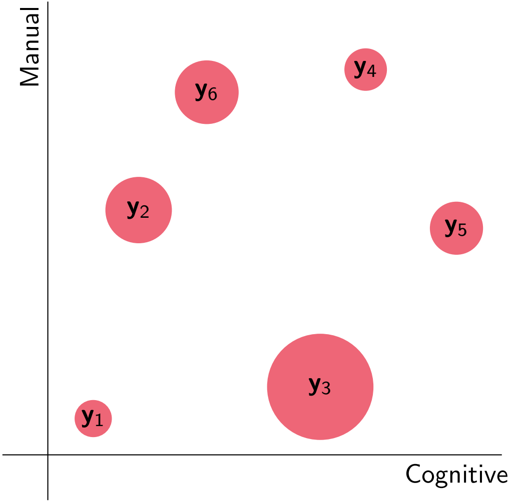
- There are several dimensions of skill

Tasks



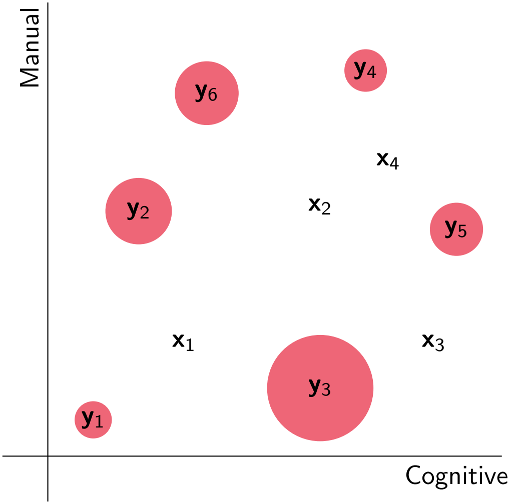
- There are several dimensions of skill
- There are K discrete tasks $\{\mathbf{y}_k\}_{k=1}^K$ that firms must complete
- Tasks are defined by their relative difficulty in each dimension of skill

Tasks



- There are several dimensions of skill
- There are K discrete tasks $\{\mathbf{y}_k\}_{k=1}^{K}$ that firms must complete
- Tasks are defined by their relative difficulty in each dimension of skill
- Tasks come in fixed proportions: distribution is given by $G(\mathbf{y})$
  I'll refer to this as the vector $\mathbf{G} \in \Delta^K$
- Firms can produce more by completing more tasks (in the same proportions). Scale distribution by a factor s

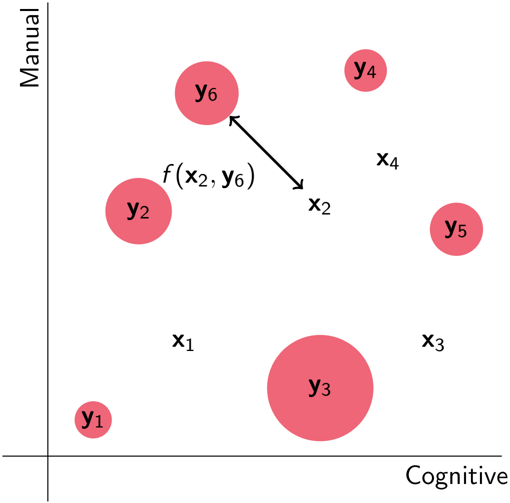
Workers



- There are a finite number of worker types $\mathbf{x}$ in the economy defined by their relative skills in each dimension

  Worker types are synonymous with occupation

Workers



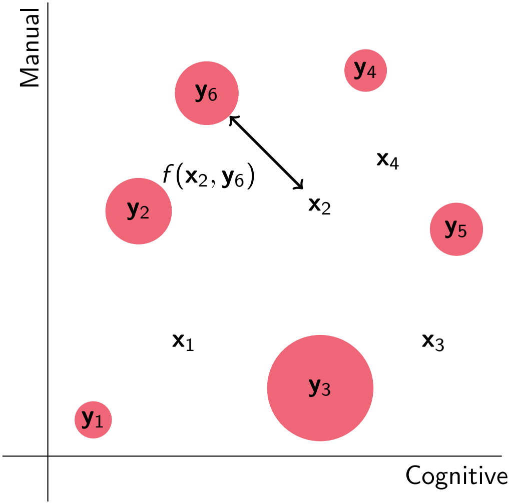
- There are a finite number of worker types $\mathbf{x}$ in the economy defined by their relative skills in each dimension

  Worker types are synonymous with occupation

- When a worker of type $\mathbf{x}$ is paired with a task $\mathbf{y}$, they produce a unit of output with quality $f(\mathbf{x}, \mathbf{y})$

- How close $\mathbf{x}$ and $\mathbf{y}$ are tells you (roughly) how well suited the worker is to do the task

Workers

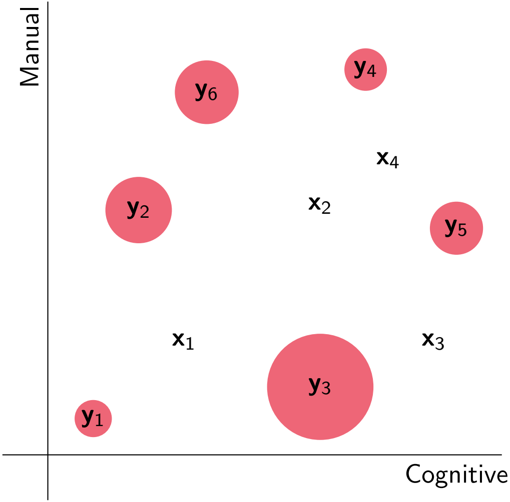


- There are a finite number of worker types $\mathbf{x}$ in the economy defined by their relative skills in each dimension

  Worker types are synonymous with occupation

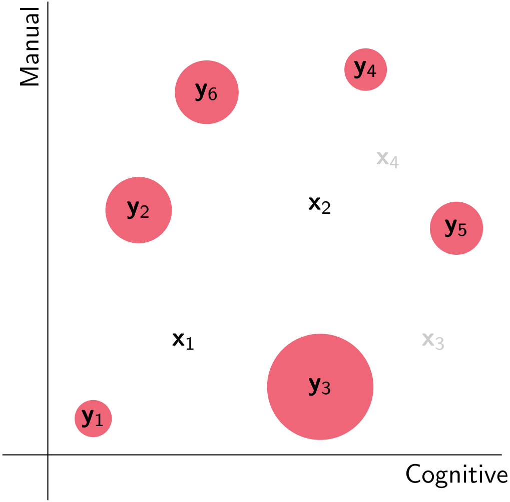
- When a worker of type $\mathbf{x}$ is paired with a task $\mathbf{y}$, they produce a unit of output with quality $f(\mathbf{x}, \mathbf{y})$
- How close $\mathbf{x}$ and $\mathbf{y}$ are tells you (roughly) how well suited the worker is to do the task
- Workers have idiosyncratic productivity ν
  - They supply units of effective labor

Firms



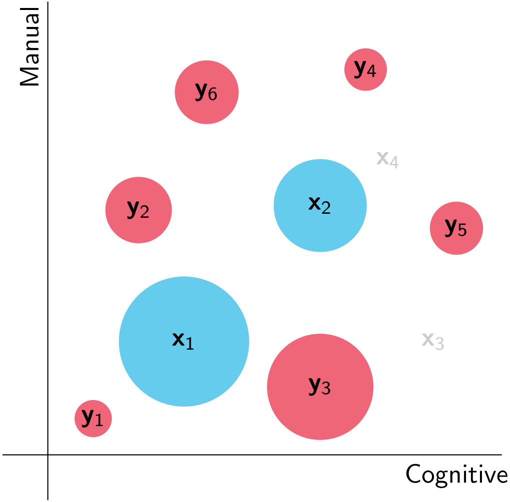
- Firms produce output by assigning tasks to workers

Firms



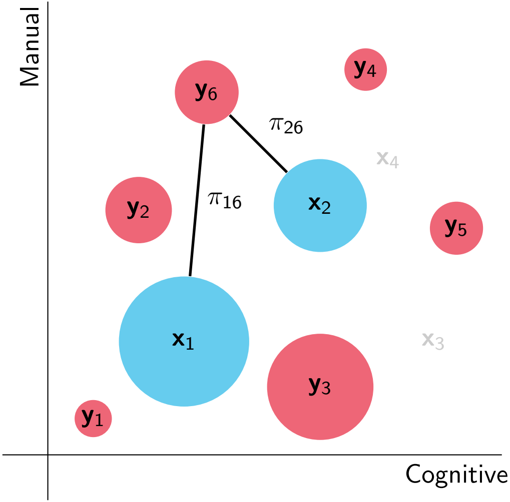
- Firms produce output by assigning tasks to workers
- Choose which set $\mathcal{X}_j$ of workers to hire
- Pay a fixed cost κ for each worker type

Firms



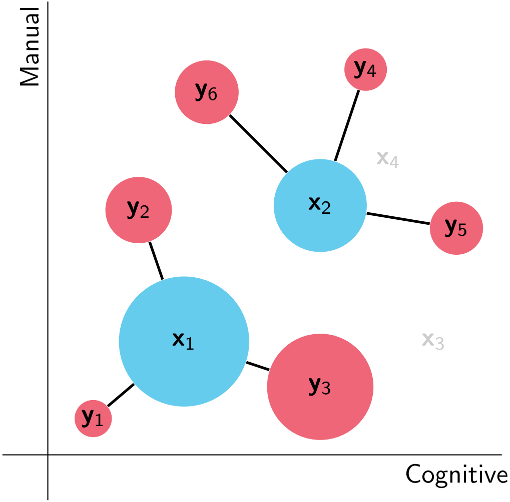
- Firms produce output by assigning tasks to workers
- Choose which set $\mathcal{X}_j$ of workers to hire
- Pay a fixed cost κ for each worker type
- Choose how much labor $\mathbf{L}_n$ to hire for workers of type $\mathbf{x}_n$

Firms



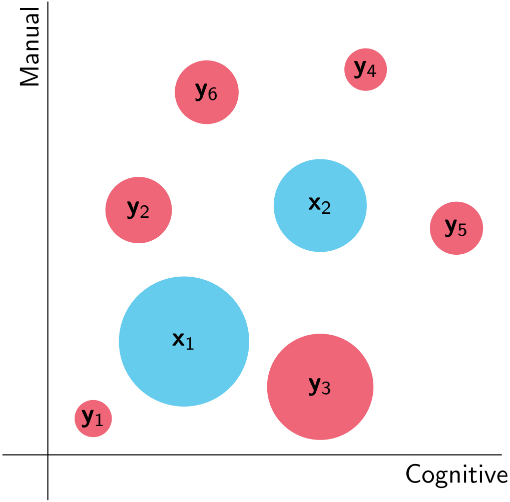
- Firms produce output by assigning tasks to workers
- Choose which set $\mathcal{X}_j$ of workers to hire
- Pay a fixed cost κ for each worker type
- Choose how much labor $\mathbf{L}_n$ to hire for workers of type $\mathbf{x}_n$
- Choose a time allocation π_{nk}: the time worker $\mathbf{x}_n$ spends on task $\mathbf{y}_k$

Firms



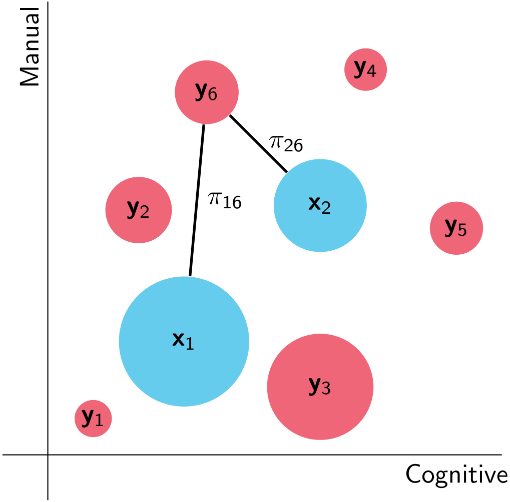
- Firms produce output by assigning tasks to workers
- Choose which set $\mathcal{X}_j$ of workers to hire
- Pay a fixed cost κ for each worker type
- Choose how much labor $\mathbf{L}_n$ to hire for workers of type $\mathbf{x}_n$
- Choose a time allocation π_{nk}: the time worker $\mathbf{x}_n$ spends on task $\mathbf{y}_k$
- If firms assign each task to a single worker, we call this a **pure assignment solution**.

Time Allocation: Feasibility



- Which time allocations are feasible?

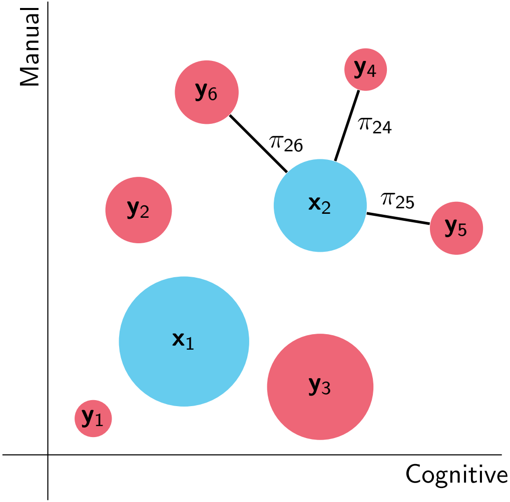
Time Allocation: Feasibility



- Which time allocations are feasible?
- Constraint 1: Every task k must be completed

$$\sum_{n=1}^{N} \pi_{nk} = \mathbf{G}_k s \tag{1}$$

Time Allocation: Feasibility



- Which time allocations are feasible?
- Constraint 1: Every task k must be completed

$$\sum_{n=1}^{N} \pi_{nk} = \mathbf{G}_k s \tag{1}$$

- Constraint 2: Each worker type n cannot be over-utilized

$$\sum_{k=1}^{K} \pi_{nk} \leq \mathbf{L}_n \tag{2}$$

Environment

- Each firm j produces a differentiated good q_j and has an idiosyncratic productivity z_j
- Final goods firm aggregates output from each individual firm:

$$Q = \left(\int q_j^{\sigma} \mathrm{d}j \right)^{\frac{1}{\sigma}} \tag{3}$$

- Each firm faces a downward sloping inverse demand curve for their product variety $p(q_j)$
- Firms must pay an occupation specific wage w_n per efficiency unit of labor for each worker type $\mathbf{x}_n$
- Firm j's output aggregates quality of worker output in each task:

$$q_j = z_j \left[\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} \right]^{\frac{1}{\eta}} \tag{4}$$

Firm Problem

1. Choose q_j given inverse demand curve, and the number of types of workers N to hire:

$$\max_{q_j, N} p(q_j)q_j - c^N(q_j, z_j) - \kappa \times N \tag{5}$$

2. Choose a set of N worker types $\mathcal{X}_j$, labor quantities $\mathbf{L}$, the time allocation π and the scale of production s to minimize costs:

$$\begin{aligned}
c^N(q_j, z_j) = \min_{\mathcal{X}_j, \mathbf{L}_j, \pi, s} \quad & \sum_{n=1}^{N} L_n w_n && \text{Total Costs} \\
\text{s.t.} \quad & \sum_{k=1}^{K} \pi_{nk} \leq \mathbf{L}_n \quad \forall n && \text{No worker over utilized} \\
& \sum_{n=1}^{N} \pi_{nk} = s \times \mathbf{G}_k \quad \forall k && \text{Every task is fully assigned} \\
& z_j \left[\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} \right]^{\frac{1}{\eta}} \geq q_j && \text{Output Constraint}
\end{aligned} \tag{6}$$

Firm Problem

1. Choose q_j given inverse demand curve, and the number of types of workers N to hire:
$$\max_{q_j, N} p(q_j)q_j - c^N(q_j, z_j) - \kappa \times N \tag{5}$$

2. Choose a set of N worker types $\mathcal{X}_j$, labor quantities $\mathbf{L}$, the time allocation π and the scale of production s to minimize costs:

$$\begin{aligned}
c^N(q_j, z_j) = \min_{\mathcal{X}_j, \mathbf{L}_j, \pi, s} \quad & \sum_{n=1}^{N} L_n w_n && \text{Total Costs} \\
\text{s.t.} \quad & \sum_{k=1}^{K} \pi_{nk} \leq \mathbf{L}_n \quad \forall n && \text{No worker over utilized} \\
& \sum_{n=1}^{N} \pi_{nk} = s \times \mathbf{G}_k \quad \forall k && \text{Every task is fully assigned} \\
& z_j \left[\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} \right]^{\frac{1}{\eta}} \geq q_j && \text{Output Constraint}
\end{aligned} \tag{6}$$

Firm Problem

1. Choose q_j given inverse demand curve, and the number of types of workers N to hire:

$$\max_{q_j, N} p(q_j)q_j - c^N(q_j, z_j) - \kappa \times N \tag{5}$$

2. Choose a set of N worker types $\mathcal{X}_j$, labor quantities $\mathbf{L}$, the time allocation π and the scale of production s to minimize costs:

$$\begin{aligned}
c^N(q_j, z_j) = \min_{\mathcal{X}_j, \mathbf{L}_j, \pi, s} \quad & \sum_{n=1}^{N} L_n w_n && \text{Total Costs} \\
\text{s.t.} \quad & \sum_{k=1}^{K} \pi_{nk} \leq \mathbf{L}_n \quad \forall n && \text{No worker over utilized} \\
& \sum_{n=1}^{N} \pi_{nk} = s \times \mathbf{G}_k \quad \forall k && \text{Every task is fully assigned} \\
& z_j \left[\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} \right]^{\frac{1}{\eta}} \geq q_j && \text{Output Constraint}
\end{aligned} \tag{6}$$

Firm Problem

1. Choose q_j given inverse demand curve, and the number of types of workers N to hire:

$$\max_{q_j, N} p(q_j)q_j - c^N(q_j, z_j) - \kappa \times N \tag{5}$$

2. Choose a set of N worker types $\mathcal{X}_j$, labor quantities $\mathbf{L}$, the time allocation π and the scale of production s to minimize costs:

$$\begin{aligned}
c^N(q_j, z_j) = \min_{\mathcal{X}_j, \mathbf{L}_j, \pi, s} \quad & \sum_{n=1}^{N} L_n w_n && \text{Total Costs} \\
\text{s.t.} \quad & \sum_{k=1}^{K} \pi_{nk} \leq \mathbf{L}_n \quad \forall n && \text{No worker over utilized} \\
& \sum_{n=1}^{N} \pi_{nk} = s \times \mathbf{G}_k \quad \forall k && \text{Every task is fully assigned} \\
& z_j \left[\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} \right]^{\frac{1}{\eta}} \geq q_j && \text{Output Constraint}
\end{aligned} \tag{6}$$

Firm Problem

1. Choose q_j given inverse demand curve, and the number of types of workers N to hire:

$$\max_{q_j, N} p(q_j)q_j - c^N(q_j, z_j) - \kappa \times N \tag{5}$$

2. Choose a set of N worker types $\mathcal{X}_j$, labor quantities $\mathbf{L}$, the time allocation π and the scale of production s to minimize costs:

$$\begin{aligned}
c^N(q_j, z_j) = \min_{\mathcal{X}_j, \mathbf{L}_j, \pi, s} \quad & \sum_{n=1}^{N} L_n w_n & \text{Total Costs} \\
\text{s.t.} \quad & \sum_{k=1}^{K} \pi_{nk} \leq \mathbf{L}_n \quad \forall n & \text{No worker over utilized} \\
& \sum_{n=1}^{N} \pi_{nk} = s \times \mathbf{G}_k \quad \forall k & \text{Every task is fully assigned} \\
& z_j \left[\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} \right]^{\frac{1}{\eta}} \geq q_j & \text{Output Constraint}
\end{aligned} \tag{6}$$

Firm Optimality Conditions

Size N firm's FOC for $\mathbf{L}_n$:

$$\underbrace{\frac{w_n \mathbf{L}_n^\star}{\sum_{i=1}^N w_i \mathbf{L}_i^\star}}_{\text{Worker } n\text{'s share of the wage bill}} \approx \underbrace{\sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^\eta \pi_{nk}^\star}_{\text{Worker } n\text{'s share of output}} = s^\star \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk}^\star \times \mathbf{G}_k \qquad (7)$$

where $\delta_{nk}^\star = \frac{\pi_{nk}^\star}{s^\star \mathbf{G}_k}$ is the share of task k assigned to worker n

- Worker n's share of the wage bill must be exactly equal to their share of total output
- Workers' marginal products are balanced against their wages
- **Key data objects:** $w_n, \mathbf{x}_n, \mathbf{L}_n$

Firm Optimality Conditions

Size N firm's FOC for $\mathbf{L}_n$:

$$\underbrace{\frac{w_n \mathbf{L}_n^\star}{\sum_{i=1}^N w_i \mathbf{L}_i^\star}}_{\text{Worker } n\text{'s share of the wage bill}} \approx \underbrace{\sum_{k=1}^K f(\mathbf{x}_n, \mathbf{y}_k)^\eta \pi_{nk}^\star}_{\text{Worker } n\text{'s share of output}} = s^\star \sum_{k=1}^K f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk}^\star \times \mathbf{G}_k \qquad (7)$$

where $\delta_{nk}^\star = \frac{\pi_{nk}^\star}{s^\star \mathbf{G}_k}$ is the share of task k assigned to worker n

- Worker n's share of the wage bill must be exactly equal to their share of total output
- Workers' marginal products are balanced against their wages
- **Key data objects:** $w_n, \mathbf{x}_n, \mathbf{L}_n$

Firm Optimality Conditions

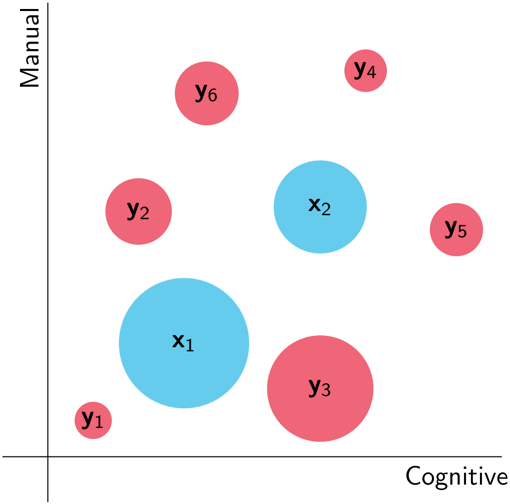
Size N firm's FOC for $\mathbf{L}_n$:

$$\underbrace{\frac{w_n \mathbf{L}_n^\star}{\sum_{i=1}^N w_i \mathbf{L}_i^\star}}_{\text{Worker } n\text{'s share of the wage bill}} \approx \underbrace{\sum_{k=1}^K f(\mathbf{x}_n, \mathbf{y}_k)^\eta \pi_{nk}^\star}_{\text{Worker } n\text{'s share of output}} = s^\star \sum_{k=1}^K f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk}^\star \times \mathbf{G}_k \tag{7}$$

where $\delta_{nk}^\star = \frac{\pi_{nk}^\star}{s^\star \mathbf{G}_k}$ is the share of task k assigned to worker n

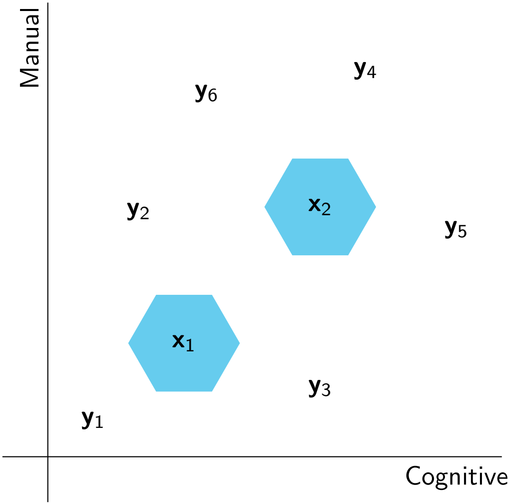
- Worker n's share of the wage bill must be exactly equal to their share of total output
- Workers' marginal products are balanced against their wages
- **Key data objects:** $w_n, \mathbf{x}_n, \mathbf{L}_n$

How much can we learn from these FOCs?



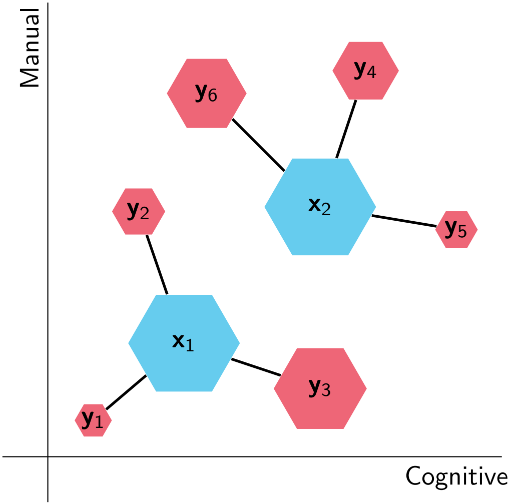
- Suppose we know f. Which task distributions $\mathbf{G}$ are consistent with the observed wage bill shares?
- Constraints are written in terms of hours spent on tasks $\mathbf{G}_k s$, and hours of labor hired $\mathbf{L}_n$

How much can we learn from these FOCs?



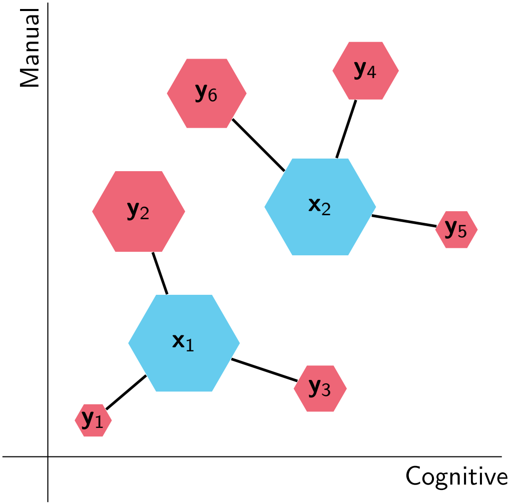
- Suppose we know f. Which task distributions $\mathbf{G}$ are consistent with the observed wage bill shares?
- Constraints are written in terms of hours spent on tasks $\mathbf{G}_k s$, and hours of labor hired $\mathbf{L}_n$
- FOCs say we need:
  1. wage bill shares $w_n \mathbf{L}_n / \sum w_i \mathbf{L}_i$

How much can we learn from these FOCs?



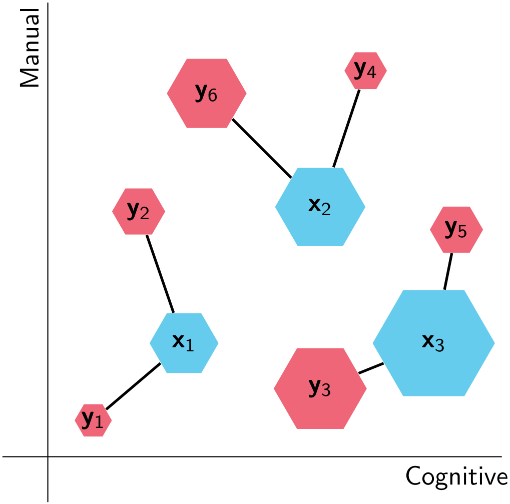
- Suppose we know f. Which task distributions $\mathbf{G}$ are consistent with the observed wage bill shares?
- Constraints are written in terms of hours spent on tasks $\mathbf{G}_k s$, and hours of labor hired $\mathbf{L}_n$
- FOCs say we need:
  1. wage bill shares $w_n \mathbf{L}_n / \sum w_i \mathbf{L}_i$
  2. worker/task output shares $f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk} \mathbf{G}_k$

How much can we learn from these FOCs?



- Suppose we know f. Which task distributions $\mathbf{G}$ are consistent with the observed wage bill shares?
- Constraints are written in terms of hours spent on tasks $\mathbf{G}_k s$, and hours of labor hired $\mathbf{L}_n$
- FOCs say we need:
    1. wage bill shares $w_n \mathbf{L}_n / \sum w_i \mathbf{L}_i$
    2. worker/task output shares $f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk} \mathbf{G}_k$
- From just two workers, not identified.

How much can we learn from these FOCs?



- Suppose we know f. Which task distributions $\mathbf{G}$ are consistent with the observed wage bill shares?
- Constraints are written in terms of hours spent on tasks $\mathbf{G}_k s$, and hours of labor hired $\mathbf{L}_n$
- FOCs say we need:
  1. wage bill shares $w_n \mathbf{L}_n / \sum w_i \mathbf{L}_i$
  2. worker/task output shares $f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \delta_{nk} \mathbf{G}_k$
- From just two workers, not identified.
- Adding firm with three pins $\mathbf{G}$ down

Identification

Size k firm's FOC for worker j:

$$\underbrace{\frac{w_n \mathbf{L}_n^\star}{\sum_{i=1}^N w_i \mathbf{L}_i^\star}}_{\text{Worker } n\text{'s share of the wage bil}} \approx \underbrace{\sum_{k=1}^K f(\mathbf{x}_n, \mathbf{y}_k)^\eta \pi_{nk}^\star}_{\text{Worker } n\text{'s share of output}} = s^\star \sum_{k=1}^K f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk}^\star \times \mathbf{G}_k \tag{8}$$

where $\delta_{nk}^\star = \frac{\pi_{nk}^\star}{s^\star \mathbf{G}_k}$ is the share of task k assigned to worker n

- **Key Assumption:** All firms within an industry have a common production function
- As firms grow larger, they are hire more types of workers and use them more effectively.
- The occupation-specific wage bill shares within the firm and *across the firm size distribution* pin down the distribution of $\mathbf{G}$
- I show in the paper: $\mathbf{G}$ is locally identified, and we are guaranteed as many degrees as freedom as the number of workers in the largest firm Theorem
- In practice: observing the same occupation with a different configuration of coworkers provides additional information about what the distribution of tasks must be

Section 4

Estimation

Parametric Assumptions and Estimation Sample

- Parametric Assumptions
  - Production function:

$$f(\mathbf{x}, \mathbf{y}) = \text{logit}^{-1}\left(x'Ay + (x-y)'B(x-y)\right)$$

  - Pick a discretization of the task space $\mathcal{X}$ that is as fine as you can (for now, $8 \times 8 \times 8$)
  - Marginal distributions of tasks are distributed Beta(α_d, β_d).
  - Joint distribution is a Gumbel Copula.

- Estimation Sample:
  - All manufacturing firms and their workers
  - Occupations at the 1 digit level

Estimation Strategy

Proceed in 3 Steps

1. Use nonlinear GMM on moment conditions implied by FOCS to recover estimates of the production function parameters A and B, and the parameters of the task distribution

2. Back out estimates of firm productivity z and output q from firms' wage bill

   ▸ Recovering q and z

3. Estimate fixed costs κ to match slope of relationship between productivity z and number of occupations k

Estimation Results: Production Function

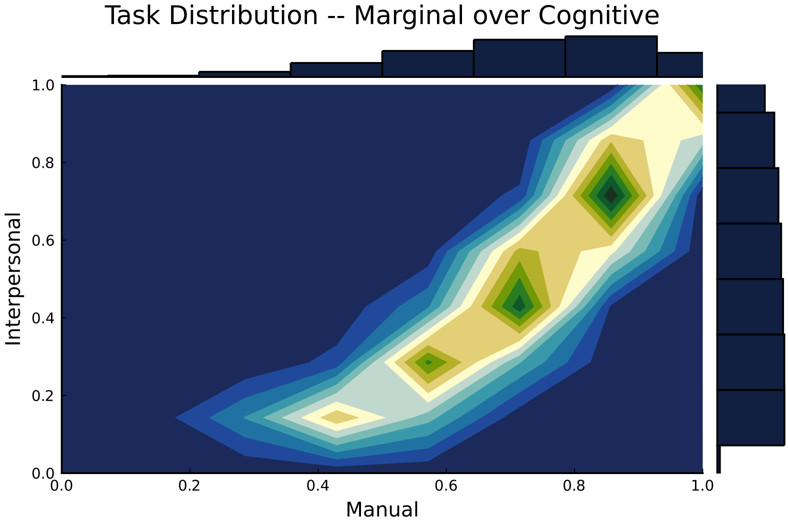
Manufacturing Industry, 2000

	Absolute Advantage			Comparative Advantage		
	Cognitive	Manual	Interpersonal	Cognitive	Manual	Interpersonal
Cognitive	1.575	-1.264	-2.161	3.499	0.687	-0.431
Manual	-8.611	2.287	4.617	0.687	0.267	0.008
Interpersonal	9.961	-0.701	-2.686	-0.431	0.008	0.123

- $\hat{\eta} = 0.922$: Firms have modestly increasing returns to scale
- $\hat{\kappa} = 20.925$: Equivalent of annual salary of 1.5 workers hired at the minimum wage

Estimation Results: Task Distribution $G(\mathbf{x})$

Manufacturing, 2000



▸ $C \times M$ ▸ $C \times I$

Decomposing Firm TFP

- Recall that we can write firm level output as:

$$q_j = \underbrace{z_j \times \overbrace{\left(\sum_{n=1}^{N}\sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} p_{jnk}\right)^{\frac{1}{\eta}}}^{\text{Endogenous Component } (\rho_j)}}_{\text{Labor Productivity}} \times L^{\frac{1}{\eta}} \tag{9}$$

where $p_{jnk} = \frac{\pi_{jnk}}{L_j}$ is the share of firm j's labor allocated to occupation n working on task k.

- What portion of the variance of "observed" labor productivity is due to the exogenous piece z_j vs. the endogenous piece ρ_j?

	Comp	Share
$\text{Var}(\log(z_j \rho_j))$	0.128	—
$\text{Var}(\log(z_j))$	0.033	25.656
$\text{Var}(\log(\rho_j))$	0.048	37.516
$2\ \text{Cov}(\log(z_j), \log(\rho_j))$	0.047	36.828

Decomposing Firm TFP

- Recall that we can write firm level output as:

$$q_j = \underbrace{z_j \times \overbrace{\left(\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^\eta p_{jnk} \right)^{\frac{1}{\eta}}}^{\text{Endogenous Component } (\rho_j)}}_{\text{Labor Productivity}} \times L^{\frac{1}{\eta}} \tag{9}$$

  where $p_{jnk} = \frac{\pi_{jnk}}{L_j}$ is the share of firm j's labor allocated to occupation n working on task k.

- What portion of the variance of "observed" labor productivity is due to the exogenous piece z_j vs. the endogenous piece ρ_j?

	Comp	Share
$\text{Var}(\log(z_j \rho_j))$	0.128	—
$\text{Var}(\log(z_j))$	0.033	25.656
$\text{Var}(\log(\rho_j))$	0.048	37.516
$2\ \text{Cov}(\log(z_j), \log(\rho_j))$	0.047	36.828

Decomposing Firm TFP

- Recall that we can write firm level output as:

$$q_j = \underbrace{z_j \times \overbrace{\left(\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} p_{jnk} \right)^{\frac{1}{\eta}}}^{\text{Endogenous Component } (\rho_j)}}_{\text{Labor Productivity}} \times L^{\frac{1}{\eta}} \tag{9}$$

where $p_{jnk} = \frac{\pi_{jnk}}{L_j}$ is the share of firm j's labor allocated to occupation n working on task k.

- What portion of the variance of "observed" labor productivity is due to the exogenous piece z_j vs. the endogenous piece ρ_j?

	Comp	Share
$\text{Var}(\log(z_j \rho_j))$	0.128	—
$\text{Var}(\log(z_j))$	0.033	25.656
$\text{Var}(\log(\rho_j))$	0.048	37.516
$2\ \text{Cov}(\log(z_j), \log(\rho_j))$	0.047	36.828

Counterfactual

- We have estimates of the task distribution $G(\mathbf{x})$, the worker-task production function $f(\mathbf{x}, \mathbf{y})$, and the fixed costs of increasing the number of occupations κ
- Three Counterfactuals:
  - Set κ to zero (all firms use the most specialized technology)
  - Set κ to a large number (all firms use the least specialized technology)
  - Double κ (intermediate)
- Rescale prices so that the total quantity of effective labor $L = \sum_{n=1}^{N} \int_0^1 \mathbf{L}_{j,n} \mathrm{d}j$ in the remains fixed. Equilibrium

Counterfactual Results

	Baseline	$\kappa = 0$	$\kappa = 2 \times \hat{\kappa}$	$\kappa =$ Large
% Δ Consumption		1.417	-0.727	-11.284
% Δ Wage		0.065	-0.046	-1.009
% Δ Output		1.203	-0.619	-9.676
Cognitive	0.374	0.369	0.375	0.370
Manual	0.431	0.463	0.420	0.321
Interpersonal	0.356	0.337	0.362	0.405

Conclusion

- Documented that larger firms hire systematically different types of workers, and spread them farther apart in the skill space
- Developed a novel task assignment model of firm production and occupational choice, and showed how to identify it using variation in firms' occupation-specific wage bill shares
- Found that 36% of the total variance of observed firm labor productivity is due to their *endogenous* choice of which types of workers to use in production
- Modest gains to subsidizing firm organizational fixed costs ($\sim$ 1.2% increase in output)
- Enormous output costs to shutting down specialization channel ($\sim$ 9.6% decrease in output)

Thank You!

Section 5

Back Matter

Section 6

Proposition 1: Linearity

Firm Hiring Problem: Linearity

Lemma 1

The firm's cost function $c^k(q; z)$ is linear in both q and z.

We can therefore rewrite the firm's problem as

$$c^k = \min_{\mathbf{y}_j, \mathbf{M}_j, \ell} \sum_{j=1}^{k} \mathbf{M}_j \ell w(\mathbf{y}_j) \quad \text{s.t. } \ell Q\left(Y, \mathbf{M}\right) \geq 1 \tag{10}$$

◂ Back

Proposition: c^n is linear in q

- That is, for all $\lambda \in \mathbb{R}$, $c^n(\lambda q; z) = \lambda c^n(q; z)$
- Let $M^\star = \sum_{i=1}^n m_i^\star \delta_{\mathbf{y}_i^\star}$ and $\mathbf{y}^\star$ solve the cost minimization problem at q.
  - Since this is feasible, we know that $zQ(\mathbf{y}^\star, M^\star) \geq q$
  - Since Q is linear in M, we know that $zQ(\mathbf{y}^\star, \lambda M^\star) = \lambda zQ(\mathbf{y}^\star, M^\star) \geq \lambda q$
  - Since this allocation is feasible, we know that $c^n(\lambda q; z) \leq \sum_{i=1}^n \lambda m_i^\star w(\mathbf{y}_i^\star) = \lambda c^n(q; z)$ by definition of c^n
- Suppose that $c^n(\lambda q; z) < \lambda c^n(q; z)$
  - Let $\hat{M} = \sum_{i=1}^n \hat{m}_i \delta_{\mathbf{y}_i^\star}$ and $\hat{\mathbf{y}}$ solve the cost minimization problem at λq
  - Note that if $zQ(\hat{\mathbf{y}}, \hat{M}) \geq \lambda q$, then by linearity of Q

$$zQ\left(\hat{\mathbf{y}}, \frac{1}{\lambda}\hat{M}\right) = \frac{1}{\lambda}\left(zQ(\hat{\mathbf{y}}, \hat{M}\right) \geq \frac{\lambda q}{\lambda} = q$$

so $(\hat{\mathbf{y}}, \hat{M})$ is feasible to produce q.

(cont...)

Back

Proposition: c^n is linear in q (cont.)

- But that means

$$\begin{aligned} c^n(q;z) &\leq \frac{1}{\lambda}\sum_{i=1}^{n} \hat{m}_i w(\hat{\mathbf{y}}_i) && \text{By optimality of } c^n \\ &= \frac{1}{\lambda} c^n(\lambda q; z) && \text{By def of } (\hat{\mathbf{y}}, \hat{M}) \\ &< \frac{1}{\lambda}\lambda c^n(q;z) && \text{By assumption} \\ &= c^n(q;z) \end{aligned}$$

  which is a contradiction.

- So $c^n(\lambda q; z) = \lambda c^n(q; z)$

□

Back

Section 7

Outsourcing Extension

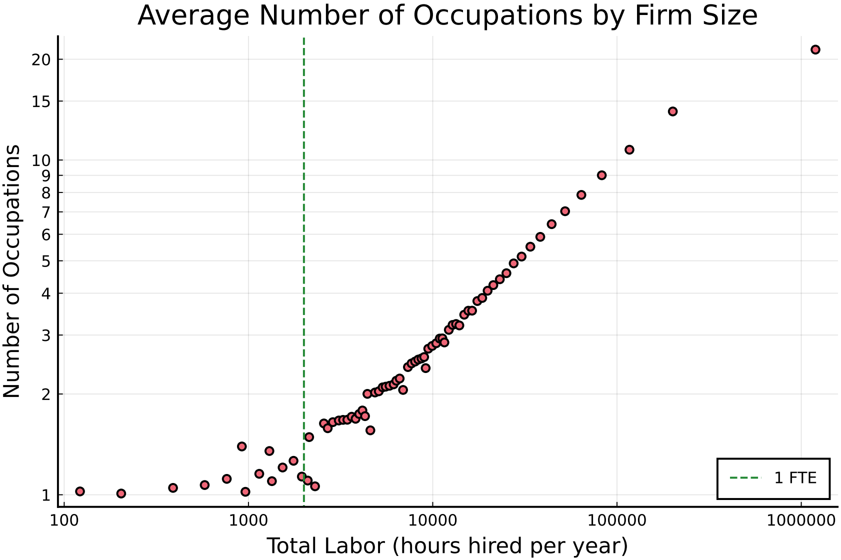
Outsourcing Extension

- For each task $\mathbf{x}_i$, firms can choose to either produce output in-house, and assign it to a worker (π_{ij}) or to contract it out (σ_i), at a cost of $\mu w(\mathbf{x}_i)$

$$
\begin{aligned}
c^k(q, \mathbf{Y}) = \min_{\pi, \sigma, M_j, s} \quad & \underbrace{\sum_{j=1}^{k} M_j w(\mathbf{y}_j)}_{\text{Wages (in-house)}} + \underbrace{\sum_{i=1}^{n} \sigma_i \mu w(\mathbf{x}_i)}_{\text{Outsourcing costs}} \\
\text{s.t.} \quad & \underbrace{\sum_{i=1}^{n} \sum_{j=1}^{k} f(\mathbf{x}_i, \mathbf{y}_j) \pi_{ij}}_{\text{In-house production}} + \underbrace{\sum_{i=1}^{n} f(\mathbf{x}_i, \mathbf{x}_i) \sigma_i}_{\text{Outsourced production}} \geq q \\
& \sum_{i=1}^{n} \pi_{ij} = \mathbf{M}_j \qquad \forall j \\
& \sum_{j=1}^{k} \pi_{ij} + \sigma_i = s \times \mathbf{G}_i \qquad \forall i
\end{aligned}
\tag{11}
$$

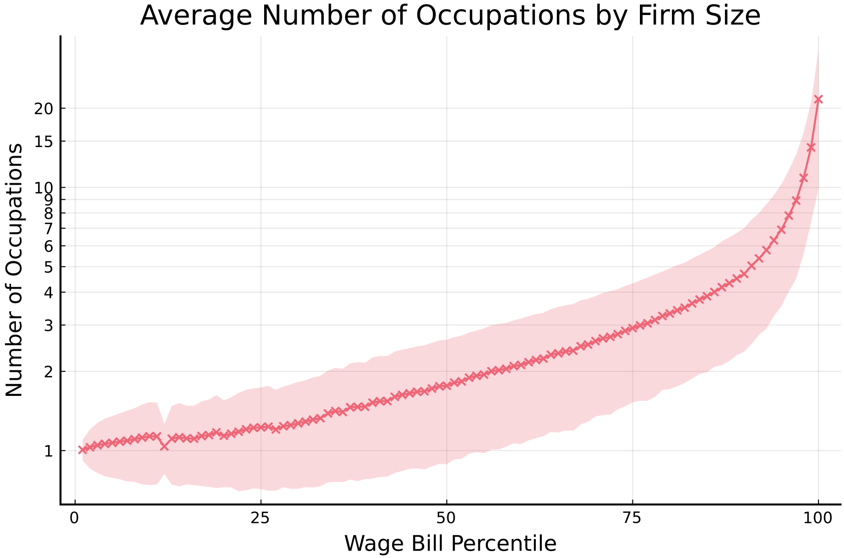
Back

Larger Firms Hire More Types of Workers: Effective Labor



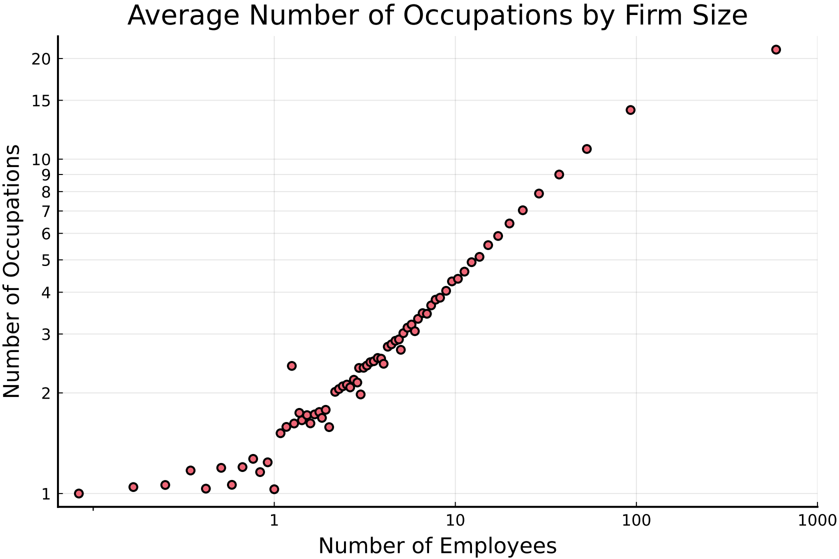
Back

Larger Firms Hire More Types of Workers: Wage Bill



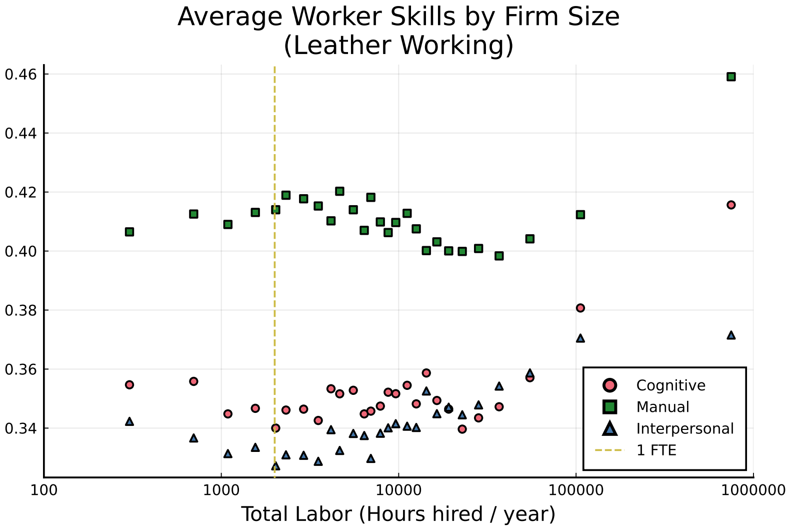
Back

Larger Firms Hire More Types of Workers: Num Employees



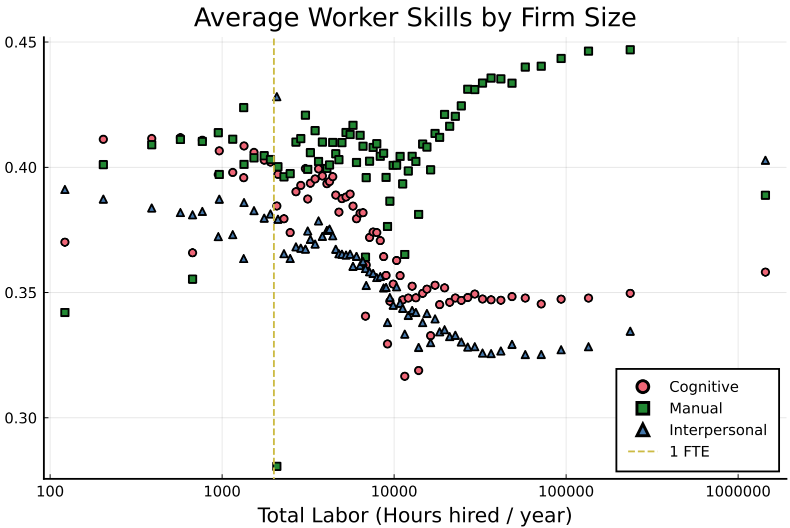
Back

Larger Firms Hire Workers with Different Skills: Leather Manufacturing



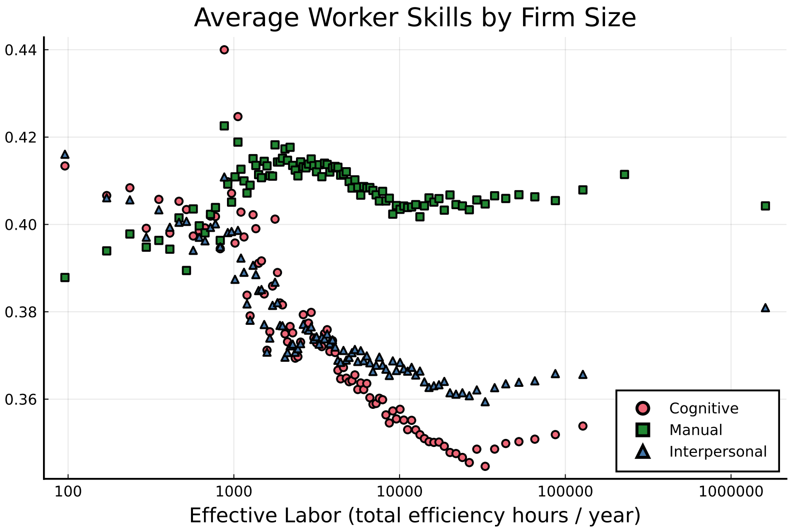
Back

Larger Firms Hire Workers with Different Skills: Total Labor



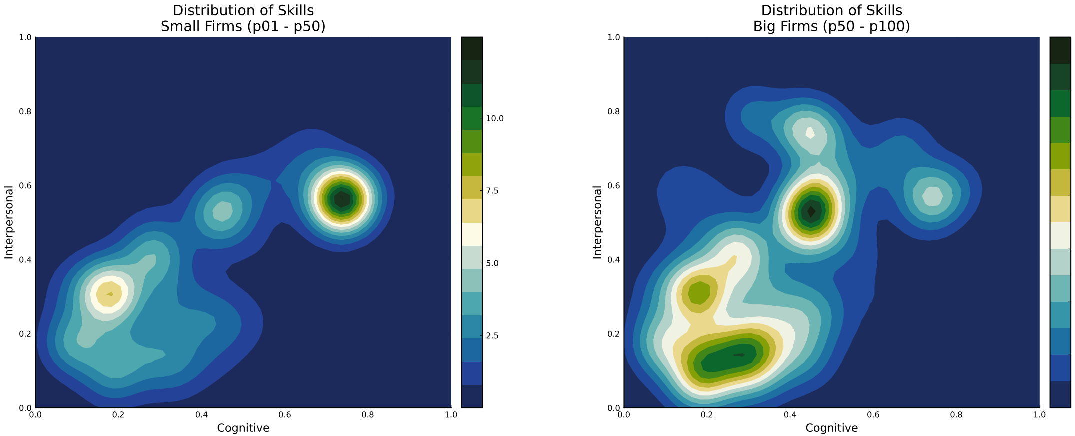
Back

Larger Firms Hire Workers with Different Skills: Industry Fixed Effects



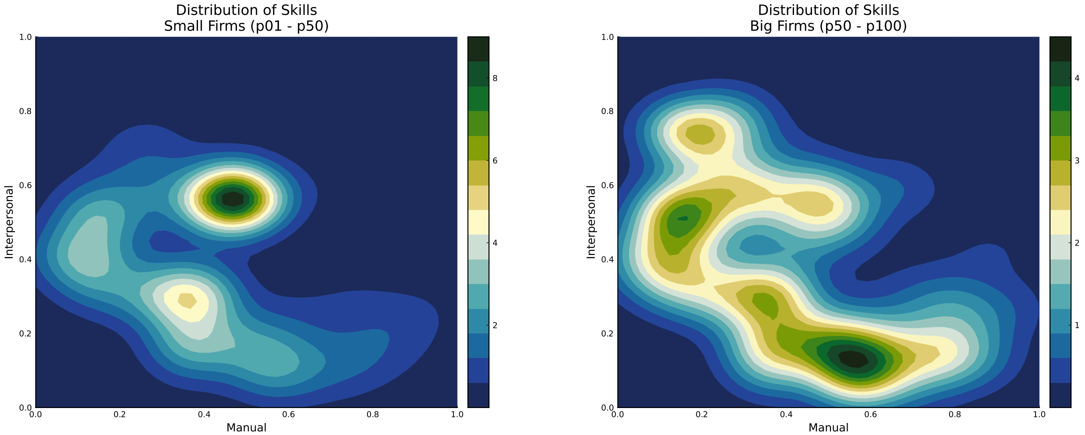
Back

Larger Firms Hire Workers with Different Skills



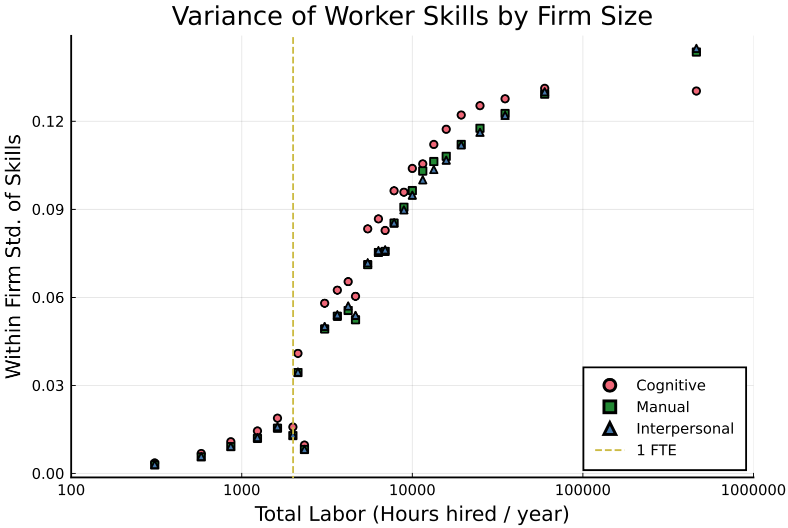
Back

Larger Firms Hire Workers with Different Skills



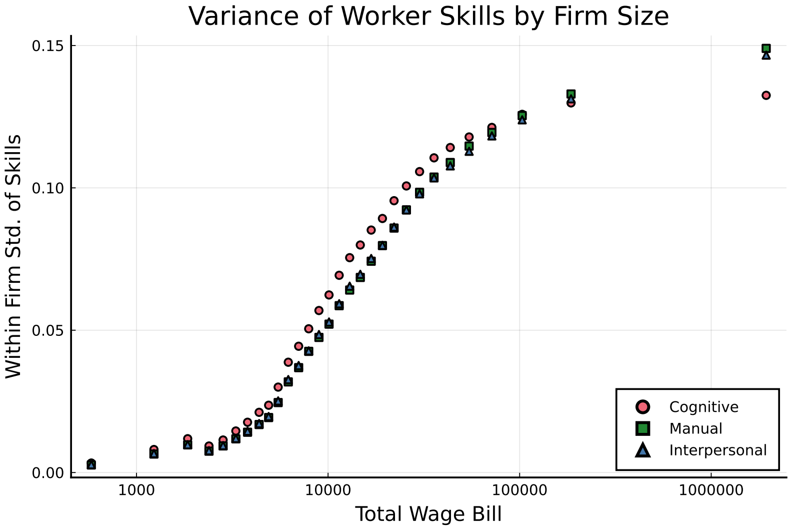
Back

Larger Firms Spread out their Workers More



Back

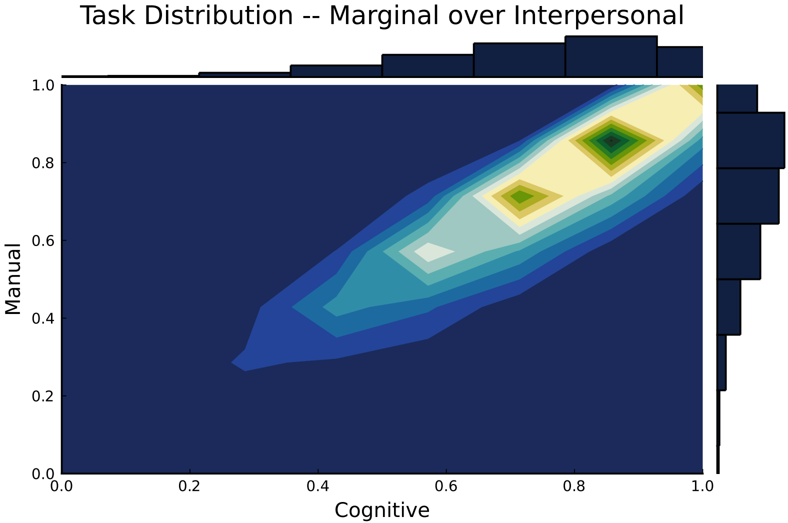
Larger Firms Spread out their Workers More



Back

Estimation Results: Task Distribution $G(\mathbf{x})$

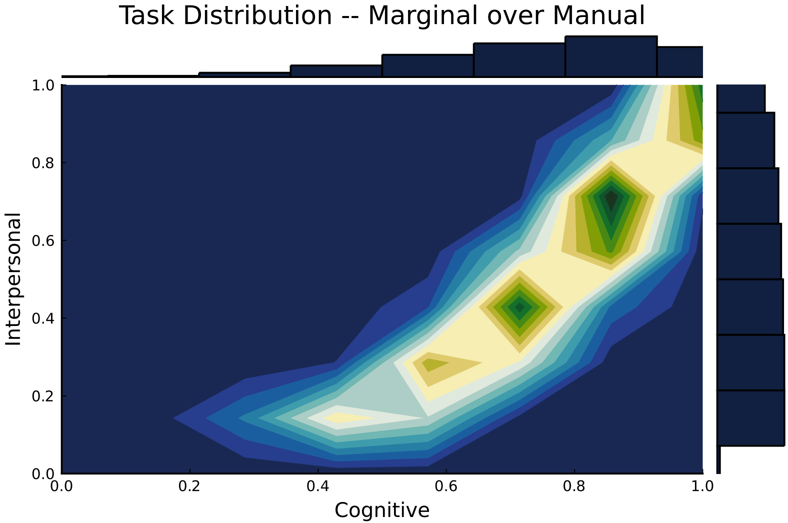
Manufacturing, 2000



Back

Estimation Results: Task Distribution $G(\mathbf{x})$

Manufacturing, 2000



Back

Recovering q and z

- I know that for each firm, their cost function is:

$$c^N(q_j, z_j) = \left(\frac{q_j}{z_j}\right)^{\eta} c_j^{\star}$$

- I impute $c_j^{\star}$ through the model, and back out $\overline{q}_j := q_j/z_j$ to match the observed total wage bill
- Given $\overline{q}$, I can back out q and z separately from the FOC of the firm problem with respect to q:

$$\begin{aligned} q_j &= \left(\frac{\eta \overline{c}_j^{\star} \overline{q}_j}{\alpha\sigma}\right)^{\frac{1}{\sigma}} \\ z_j &= \frac{1}{\overline{q}_j}\left(\frac{\eta \overline{c}_j^{\star} \overline{q}_j}{\alpha\sigma}\right)^{\frac{1}{\sigma}} \end{aligned} \tag{12}$$

Back

Narrowly Defined Industries

- Choose several narrowly defined industries (following Foster, Haltiwanger, and Syverson, 2016):
  - Sugar: "Cultivation of Sugar Cane," (01139) "Sugar Mills," (15610) and "Sugar Refining and Milling" (15628)
  - Plywood: "Manufacture of Laminated Wood and Plywood, Pressed or Agglomerated Sheets" (20214)
  - Cement Manufacturing: "Cement Manufacturing" (26204)
  - Coffee: "Coffee Growing" (01325) and "Coffee Roasting and Grinding" (15717)
- Industries that produce a homogenous commodity good: unlikely that large and small firms have systematically different production processes
- Rerun analysis industry by industry:

$$y_i = \beta_0 + \sum_{s=2}^{10} \beta^s D_i^s + \epsilon_i$$

Back

Number of Occupations

	log(Occupations)			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.042*	0.096***	0.059***	0.051***
	(0.024)	(0.030)	(0.009)	(0.019)
deciles: 2	0.169***	0.313***	0.081***	0.102***
	(0.055)	(0.054)	(0.015)	(0.037)
deciles: 3	0.531***	0.484***	0.164***	0.175***
	(0.081)	(0.063)	(0.018)	(0.044)
deciles: 4	0.888***	0.766***	0.277***	0.404***
	(0.083)	(0.064)	(0.020)	(0.046)
deciles: 5	1.433***	0.928***	0.406***	0.566***
	(0.107)	(0.065)	(0.022)	(0.053)
deciles: 6	2.306***	1.184***	0.519***	0.749***
	(0.095)	(0.064)	(0.023)	(0.048)
deciles: 7	2.862***	1.297***	0.665***	1.075***
	(0.099)	(0.068)	(0.023)	(0.049)
deciles: 8	3.023***	1.573***	0.849***	1.283***
	(0.089)	(0.065)	(0.024)	(0.050)
deciles: 9	3.354***	1.990***	1.094***	1.575***
	(0.056)	(0.065)	(0.024)	(0.048)
deciles: 10	3.533***	2.391***	1.819***	2.278***
	(0.064)	(0.088)	(0.032)	(0.063)
N	497	1061	5506	941
R^2	0.857	0.595	0.553	0.732

Back

Dispersion of Cognitive Skills

	Cognitive w/in Firm Std			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.006*	0.009***	0.007***	0.009**
	(0.004)	(0.003)	(0.001)	(0.004)
deciles: 2	0.026***	0.035***	0.012***	0.009
	(0.009)	(0.007)	(0.003)	(0.007)
deciles: 3	0.059***	0.037***	0.021***	0.027***
	(0.010)	(0.007)	(0.003)	(0.008)
deciles: 4	0.101***	0.052***	0.030***	0.069***
	(0.010)	(0.007)	(0.003)	(0.010)
deciles: 5	0.129***	0.045***	0.042***	0.087***
	(0.010)	(0.006)	(0.003)	(0.011)
deciles: 6	0.149***	0.063***	0.054***	0.119***
	(0.006)	(0.007)	(0.004)	(0.010)
deciles: 7	0.136***	0.056***	0.062***	0.153***
	(0.006)	(0.006)	(0.004)	(0.009)
deciles: 8	0.147***	0.058***	0.071***	0.150***
	(0.006)	(0.005)	(0.004)	(0.009)
deciles: 9	0.149***	0.065***	0.082***	0.163***
	(0.005)	(0.005)	(0.003)	(0.008)
deciles: 10	0.149***	0.074***	0.104***	0.175***
	(0.005)	(0.005)	(0.003)	(0.007)
N	497	1061	5506	941
R^2	0.557	0.132	0.181	0.428

◂ Back

Dispersion of Manual Skills

	Manual w/in Firm Std			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.007	0.010***	0.009***	0.009**
	(0.004)	(0.004)	(0.002)	(0.004)
deciles: 2	0.026**	0.047***	0.014***	0.014**
	(0.010)	(0.009)	(0.003)	(0.007)
deciles: 3	0.093***	0.056***	0.024***	0.026***
	(0.016)	(0.009)	(0.003)	(0.008)
deciles: 4	0.129***	0.073***	0.040***	0.058***
	(0.014)	(0.008)	(0.004)	(0.009)
deciles: 5	0.154***	0.070***	0.049***	0.059***
	(0.011)	(0.007)	(0.004)	(0.008)
deciles: 6	0.193***	0.080***	0.062***	0.078***
	(0.009)	(0.007)	(0.004)	(0.008)
deciles: 7	0.204***	0.073***	0.079***	0.108***
	(0.010)	(0.006)	(0.004)	(0.008)
deciles: 8	0.203***	0.072***	0.089***	0.120***
	(0.008)	(0.006)	(0.003)	(0.008)
deciles: 9	0.211***	0.088***	0.106***	0.137***
	(0.006)	(0.006)	(0.003)	(0.006)
deciles: 10	0.208***	0.096***	0.136***	0.157***
	(0.007)	(0.006)	(0.003)	(0.007)
N	497	1061	5506	941
R^2	0.584	0.148	0.268	0.399

Back

Dispersion of Interpersonal Skills

	Interpersonal w/in Firm Std			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.005	0.011***	0.009***	0.007**
	(0.004)	(0.004)	(0.001)	(0.003)
deciles: 2	0.020**	0.044***	0.012***	0.015**
	(0.008)	(0.009)	(0.003)	(0.006)
deciles: 3	0.053***	0.052***	0.021***	0.023***
	(0.011)	(0.009)	(0.003)	(0.007)
deciles: 4	0.091***	0.071***	0.030***	0.064***
	(0.011)	(0.008)	(0.003)	(0.009)
deciles: 5	0.103***	0.069***	0.043***	0.081***
	(0.009)	(0.008)	(0.003)	(0.009)
deciles: 6	0.141***	0.090***	0.057***	0.115***
	(0.007)	(0.008)	(0.003)	(0.008)
deciles: 7	0.146***	0.082***	0.066***	0.134***
	(0.007)	(0.007)	(0.003)	(0.008)
deciles: 8	0.149***	0.093***	0.077***	0.136***
	(0.007)	(0.006)	(0.003)	(0.007)
deciles: 9	0.145***	0.103***	0.088***	0.141***
	(0.006)	(0.006)	(0.003)	(0.006)
deciles: 10	0.143***	0.116***	0.114***	0.156***
	(0.006)	(0.006)	(0.003)	(0.005)
N	497	1061	5506	941
R^2	0.542	0.201	0.242	0.448

Back

Level of Cognitive Skills

	Cognitive Skills			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.366***	0.350***	0.340***	0.396***
	(0.027)	(0.012)	(0.008)	(0.022)
deciles: 2	0.045	-0.031**	-0.013	-0.040
	(0.036)	(0.015)	(0.010)	(0.029)
deciles: 3	0.036	-0.050***	-0.039***	0.001
	(0.035)	(0.015)	(0.010)	(0.032)
deciles: 4	0.006	-0.038***	-0.047***	-0.039
	(0.032)	(0.014)	(0.009)	(0.028)
deciles: 5	0.012	-0.064***	-0.045***	-0.028
	(0.031)	(0.014)	(0.009)	(0.029)
deciles: 6	0.004	-0.059***	-0.050***	-0.007
	(0.029)	(0.014)	(0.009)	(0.027)
deciles: 7	-0.002	-0.067***	-0.045***	-0.004
	(0.030)	(0.013)	(0.009)	(0.027)
deciles: 8	-5.496e-04	-0.066***	-0.047***	-0.015
	(0.029)	(0.013)	(0.009)	(0.027)
deciles: 9	-0.007	-0.068***	-0.044***	0.038
	(0.028)	(0.013)	(0.008)	(0.026)
deciles: 10	-0.021	-0.066***	-0.027***	-0.004
	(0.028)	(0.013)	(0.008)	(0.025)
N	497	1061	5506	941
R^2	0.025	0.080	0.016	0.017

Back

Level of Manual Skills

	Manual Skills			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.428***	0.584***	0.524***	0.356***
	(0.028)	(0.018)	(0.008)	(0.020)
deciles: 2	0.083*	-0.003	0.042***	0.053*
	(0.042)	(0.025)	(0.011)	(0.029)
deciles: 3	0.046	0.019	0.045***	0.097***
	(0.039)	(0.022)	(0.011)	(0.027)
deciles: 4	0.045	0.032	0.055***	0.096***
	(0.036)	(0.022)	(0.011)	(0.024)
deciles: 5	0.071**	0.030	0.058***	0.065***
	(0.035)	(0.020)	(0.010)	(0.024)
deciles: 6	0.074**	0.022	0.064***	0.065***
	(0.030)	(0.021)	(0.010)	(0.024)
deciles: 7	0.099***	0.048**	0.069***	0.056**
	(0.029)	(0.020)	(0.010)	(0.025)
deciles: 8	0.121***	0.045**	0.072***	0.066***
	(0.030)	(0.020)	(0.009)	(0.023)
deciles: 9	0.114***	0.043**	0.082***	0.043*
	(0.029)	(0.019)	(0.009)	(0.024)
deciles: 10	0.118***	0.034*	0.082***	0.005
	(0.028)	(0.019)	(0.009)	(0.022)
N	497	1061	5506	941
R^2	0.066	0.021	0.024	0.041

Back

Level of Interpersonal Skills

	Interpersonal Skills			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.396***	0.216***	0.288***	0.391***
	(0.024)	(0.020)	(0.008)	(0.018)
deciles: 2	-0.006	-0.033	-0.026**	-0.041
	(0.032)	(0.025)	(0.010)	(0.025)
deciles: 3	2.131e-04	-0.063***	-0.044***	-0.039
	(0.033)	(0.022)	(0.010)	(0.025)
deciles: 4	-0.030	-0.067***	-0.057***	-0.069***
	(0.029)	(0.022)	(0.009)	(0.025)
deciles: 5	-0.031	-0.078***	-0.060***	-0.040
	(0.028)	(0.021)	(0.009)	(0.025)
deciles: 6	-0.053**	-0.071***	-0.069***	-0.028
	(0.026)	(0.022)	(0.009)	(0.023)
deciles: 7	-0.083***	-0.091***	-0.064***	-0.015
	(0.025)	(0.020)	(0.009)	(0.021)
deciles: 8	-0.094***	-0.095***	-0.070***	-0.032
	(0.026)	(0.021)	(0.009)	(0.022)
deciles: 9	-0.097***	-0.094***	-0.069***	0.021
	(0.025)	(0.021)	(0.008)	(0.020)
deciles: 10	-0.105***	-0.093***	-0.060***	0.001
	(0.025)	(0.021)	(0.008)	(0.020)
N	497	1061	5506	941
R^2	0.118	0.070	0.032	0.031

Back

Theorem 2

Suppose we observe firms that hire up to N occupations in the data, and we observe data on their skills and wages. Suppose further that the function f is known, and that f distinguishes workers almost everywhere. If the number of tasks is K, and $K \leq N$ then the distribution of tasks $\mathbf{G}$ is identified within a neighborhood of the solution.

Sketch of the Proof:

- Key FOC:

$$\frac{w_n \mathbf{L}_n^\star}{\sum_{i=1}^N w_i \mathbf{L}_i^\star} \approx s^\star \sum_{k=1}^K f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk}^\star \times \mathbf{G}_k$$

  where $\delta_{nk}^\star = \frac{\pi_{nk}^\star}{s^\star \mathbf{G}_k}$ is the share of task k assigned to worker n

- Within a neighborhood of the true solution, the task assignment δ_{nk} does not change
- So for a fixed f and set of workers $\mathcal{X}_j$, this is just a linear system of equations
- Identification is equivalent to proving that the system has full row rank
- This follows from the fact that the firm chooses a pure assignment solution

Back

Equilibrium Specification

- Need to know how wages adjust to changes in firm demand for labor
- Assume: each worker is endowed with L units of labor and idiosyncratic productivity ν.
- Workers maximize:

$$\max_{n \in 1,\ldots,N} \log(w_n \nu L) - c_n \quad (13)$$

where c_n is a disamenity cost of working in a particular occupation.

- An **equilibrium** is a set of wages and quantities such that
  1. Workers are indifferent between all of the occupations (which are chosen in equilibrium)
  2. The total quantity of labor demanded, integrating across all the occupations, equals the total supply.
- Worker indifference requires that for any occupations n and n':

$$\log(w_n) - \log(w_{n'}) = c_n - c_{n'}$$

so the relative wages $w_n/w_{n'}$ are pinned down by the difference in occupation specific amenities

Back

Equilibrium Specification

- Need to know how wages adjust to changes in firm demand for labor
- Assume: each worker is endowed with L units of labor and idiosyncratic productivity ν.
- Workers maximize:

$$\max_{n \in 1,\ldots,N} \log(w_n \nu L) - c_n \tag{13}$$

where c_n is a disamenity cost of working in a particular occupation.

- An **equilibrium** is a set of wages and quantities such that
  1. Workers are indifferent between all of the occupations (which are chosen in equilibrium)
  2. The total quantity of labor demanded, integrating across all the occupations, equals the total supply.
- Worker indifference requires that for any occupations n and n':

$$\log(w_n) - \log(w_{n'}) = c_n - c_{n'}$$

so the relative wages $w_n / w_{n'}$ are pinned down by the difference in occupation specific amenities

Back

Equilibrium Specification

- Need to know how wages adjust to changes in firm demand for labor
- Assume: each worker is endowed with L units of labor and idiosyncratic productivity ν.
- Workers maximize:

$$\max_{n \in 1,\ldots,N} \log(w_n \nu L) - c_n \tag{13}$$

  where c_n is a disamenity cost of working in a particular occupation.
- An **equilibrium** is a set of wages and quantities such that
  1. Workers are indifferent between all of the occupations (which are chosen in equilibrium)
  2. The total quantity of labor demanded, integrating across all the occupations, equals the total supply.
- Worker indifference requires that for any occupations n and n':

$$\log(w_n) - \log(w_{n'}) = c_n - c_{n'}$$

  so the relative wages $w_n/w_{n'}$ are pinned down by the difference in occupation specific amenities

Back

Definition of Effective Labor

- For each worker i, working in occupation n_i and employed at firm j_i, let their monthly earnings be w_i
- For every occupation n define $\overline{w}_n$ as the average monthly earnings for workers in occupation n, weighted by total hours worked l_i
- Define worker i's effective labor supplied as

$$M_i := l_i \times \left(\frac{w_i}{\overline{w}_{n_i}} \right)$$

- Adjust total hours worked by the ratio of i's wages to their occupation's average wages

Back

References

Acemoglu, D. and D. Autor (2011): "Skills, technologies: Implications for employment and earnings," in *Handbook of labor economics*, Elsevier, vol. 4, 1043–1171.

Acemoglu, D. and P. Restrepo (2021): "Tasks, Automation, and the Rise in US Wage Inequality," Tech. rep., National Bureau of Economic Research.

Ales, L., C. Combemale, E. Fuchs, and K. Whitefoot (2021): "How It's Made: A General Theory of the Labor Implications of Technological Change," Tech. rep.

Autor, D. H. and D. Dorn (2013): "The growth of low-skill service jobs and the polarization of the US labor market," *American Economic Review*, 103, 1553–97.

Buera, F. J., J. P. Kaboski, and R. Rogerson (2015): "Skill biased structural change," Tech. rep., National Bureau of Economic Research.

Caliendo, L., F. Monte, and E. Rossi-Hansberg (2015): "The anatomy of French production hierarchies," *Journal of Political Economy*, 123, 809–852.

Caliendo, L. and E. Rossi-Hansberg (2012): "The Impact of Trade on Organization and Productivity," *The Quarterly Journal of Economics*, 127, 1393–1467.

References (cont.)

De Souza, G. (2020): "The Labor Market Consequences of Appropriate Technology," .

Foster, L., J. Haltiwanger, and C. Syverson (2016): "The slow growth of new plants: Learning about demand?" *Economica*, 83, 91–129.

Garicano, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of political economy*, 108, 874–904.

Heckman, J. J. and G. Sedlacek (1985): "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy*, 93, 1077–1125.

Katz, L. F. and K. M. Murphy (1992): "Changes in Relative Wages, 1963–1987: Supply and Demand Factors*," *The Quarterly Journal of Economics*, 107, 35–78.

Lindenlaub, I. (2017): "Sorting multidimensional types: Theory and application," *The Review of Economic Studies*, 84, 718–789.

Lise, J. and F. Postel-Vinay (2020): "Multidimensional skills, sorting, and human capital accumulation," *American Economic Review*, 110, 2328–76.

References (cont.)

OCAMPO, S. (2019): "A task-based theory of occupations with multidimensional heterogeneity," JMP.

ROSEN, S. (1978): "Substitution and division of labour," *Economica*, 45, 235–250.

——— (1982): "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, 13, 311–323.