For Whom the Bot Tolls: Skills and the Earnings Effects of AI

Lukas Freund
Columbia University

Lukas Mann
Minneapolis Fed

Columbia University
May 6, 2025

The views expressed in this presentation are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

AI capabilities are rapidly improving relative to humans

Test scores of AI systems on various capabilities relative to human performance

Our World in Data

Within each domain, the initial performance of the AI is set to −100. Human performance is used as a baseline, set to zero. When the AI's performance crosses the zero line, it scored more points than humans.

Data source: Kiela et al. (2023)

OurWorldinData.org/artificial-intelligence | CC BY

Note: For each capability, the first year always shows a baseline of −100, even if better performance was recorded later that year.

Large-scale automation *exposure*

ASK FT

FINANCIAL TIMES

HOME WORLD US COMPANIES TECH MARKETS CLIMATE OPINION LEX WORK & CAREERS LIFE & ARTS HTSI

Artificial intelligence + Add to myFT

Higher earners face greater AI exposure, study finds

Research estimates how much the fast-evolving technology hits various jobs, from software engineers to mechanics

Share

Save

Jobs most affected included blockchain engineers, clinical data managers, public relations specialists and financial quantitative analysts © Getty Images

Michael Peel in London

Published JUN 20 2024

34

BROOKINGS U.S. Trade Policy | International Affairs | U.S. Government & Politics | Technology & Information | Race in Public Policy | Topics | Regions

Generative AI, the American worker, and the future of work

Molly Kinder, Xavier de Sousa Briggs, Mark Muro, and Siliencia
October 10, 2024

- Exciting generative AI technology already has the potential to significantly disrupt a wide range of jobs. We find that 30 million U.S. workers could see at least 50% of their occupations reshaped by generative AI.
- Unlike previous automation technologies that primarily affected routine, blue-collar work, generative AI is likely to disrupt a different basis, especially in middle- to higher-paid profes
- Despite the high stakes for workers, we are very opportunities that generative AI presents joint
- The report emphasizes the importance of deven design and implementation, enhancing worker e and developing public policies that ensure work harms such as job loss and inequality

POLITICO

War in Ukraine Romania election Newsletters Podcasts Poll of Polls Policy news Events

NEWS > TECHNOLOGY

IMF report: 40 percent of jobs exposed to AI

SHARE

The report landed on Sunday, right before the World Economic Forum (WEF) in Davos, where the rise of artificial intelligence, and generative AI in particular, will be a major talking point | Stefani Reynolds/AFP via Getty Images

JANUARY 15, 2024 3:53 PM CET
BY PIETER HAECK

What will be the implications for earnings & their distribution?

- **Q: Who will gain and who will lose from AI-induced task automation?**

Is exposure good or bad?

Law example

ars TECHNICA

AI | BIZ & IT | CARS | CULTURE | GAMING | HEALTH | POLICY | SCIENCE | SECURITY | SPACE | TECH | FORUM | SUBSCRIBE

RIDING THE DIGITAL WAVES

Will the future of software development run on vibes?

Accepting AI-written code without understanding how it works is growing in popularity.

BENJ EDWARDS – 5 MAR 2025 18:41 | 285

Credit: Henrik5000 via Getty Images

CODEX

What will be the implications for earnings & their distribution?

- **Q: Who will gain and who will lose from AI-induced task automation?**

- **Intuition:** earnings effects more likely to be *negative* if a worker is specialized in tasks AI can do, *positive* if automated task was their comparative disadvantage

What will be the implications for earnings & their distribution?

- **Q: Who will gain and who will lose from AI-induced task automation?**

- **Intuition:** earnings effects more likely to be *negative* if a worker is specialized in tasks AI can do, *positive* if automated task was their comparative disadvantage

- So to answer the Q, we **need** both

  1. forward-looking measures of **task automation exposure**

     ✓vast literature *[e.g. Webb, 2020; Felten et al., 2021; Eloundou et al., 2023; ...]*

  2. **structural model** of how task exposure & skills shape earnings

- **This paper:** structural model mapped to automation exposure measures
  → **predict worker-level earnings effects of AI automation**

What we do: theory-guided measurement & counterfactual analysis

1. **Theory**: task-based model of production + occupational choice

2. **Measurement**: distribution of task-specific skills

3. **Quantitative counterfactuals** combining model + task exposure measures

What we do: theory-guided measurement & counterfactual analysis

1. **Theory**: task-based model of production + occupational choice
   - occupations bundle tasks with heterogeneous weights
   - workers have heterogeneous portfolios of skills for each task
   - choose occupation, earn wage
2. **Measurement**: distribution of task-specific skills
3. **Quantitative counterfactuals** combining model + task exposure measures

What we do: theory-guided measurement & counterfactual analysis

1. **Theory**: task-based model of production + occupational choice
   - occupations bundle tasks with heterogeneous weights
   - workers have heterogeneous portfolios of skills for each task
   - choose occupation, earn wage
2. **Measurement**: distribution of task-specific skills
   - LLMs: occupational task weights for 30 tasks (clustering of $\sim$ 20,000 O*NET tasks)
   - NLSY: worker panel of occ. choices & wages
   - estimate skill distribution using model structure
3. **Quantitative counterfactuals** combining model + task exposure measures

What we do: theory-guided measurement & counterfactual analysis

1. **Theory**: task-based model of production + occupational choice
   - occupations bundle tasks with heterogeneous weights
   - workers have heterogeneous portfolios of skills for each task
   - choose occupation, earn wage
2. **Measurement**: distribution of task-specific skills
   - LLMs: occupational task weights for 30 tasks (clustering of $\sim$ 20,000 O*NET tasks)
   - NLSY: worker panel of occ. choices & wages
   - estimate skill distribution using model structure
3. **Quantitative counterfactuals** combining model + task exposure measures
   - historical validation: **industrial robots**
   - **AI**: automation of information-processing tasks

What's new?

- **AI: measurement of job exposure**
  [Brynjolfsson et al., 2018; Webb, 2019; Felten et al., 2021; Eloundou et al., 2023; ...]
  $\Rightarrow$ map to **structural** model $\rightarrow$ individual **earnings effects** as a function of skills

- **Task-based framework** *[Acemoglu-Autor, 2011; Acemoglu-Restrepo, 2022; Freund, 2024; ...]*
  $\Rightarrow$ empirically operationalize $\rightarrow$ link to **forward-looking** automation measures

- **Multi-dimensional skills**
  [Lindenlaub, 2017; Guvenen et al., 2020; Lise-PostelVinay, 2021; Deming, 2023; Grigsby, 2023]
  $\Rightarrow$ **estimate** distribution of high-dim. task-specific skills $\rightarrow$ **skill specialization**

- **Applications of LLMs in economics research** *[Korinek, 2023; Athey et al., 2024; Dell, 2024]*
  $\Rightarrow$ use LLMs for clustering & time-share measurement $\rightarrow$ **flexible** tool

Theory

Big picture: requirements $\rightarrow$ key model features

Q: Who will gain and who will lose from AI-induced task automation?

1. Technological change as task automation *[Acemoglu-Autor, 2011; Acemoglu-Restrepo, 2018; ...]*

   $\Rightarrow$ **task-based framework**

Big picture: requirements $\rightarrow$ key model features

Q: Who will gain and who will lose from AI-induced task automation?

1. Technological change as task automation *[Acemoglu-Autor, 2011; Acemoglu-Restrepo, 2018; ...]*

   $\Rightarrow$ **task-based framework**

2. Exposure to automation varies across tasks and hence occupations *[e.g. Webb, 2020]*

   $\Rightarrow$ explicit model of **occupation-specific production technology**

Big picture: requirements $\rightarrow$ key model features

Q: Who will gain and who will lose from AI-induced task automation?

1. Technological change as task automation *[Acemoglu-Autor, 2011; Acemoglu-Restrepo, 2018; ...]*
   $\Rightarrow$ **task-based framework**
2. Exposure to automation varies across tasks and hence occupations *[e.g. Webb, 2020]*
   $\Rightarrow$ explicit model of **occupation-specific production technology**
3. Earnings effects from automation vary depending on skill portfolio *[e.g. Autor, 2015]*
   $\Rightarrow$ **high-dimensional, task-specific skills**

Big picture: requirements → key model features

Q: Who will gain and who will lose from AI-induced task automation?

1. Technological change as task automation *[Acemoglu-Autor, 2011; Acemoglu-Restrepo, 2018; ...]*
   ⇒ **task-based framework**
2. Exposure to automation varies across tasks and hence occupations *[e.g. Webb, 2020]*
   ⇒ explicit model of **occupation-specific production technology**
3. Earnings effects from automation vary depending on skill portfolio *[e.g. Autor, 2015]*
   ⇒ **high-dimensional, task-specific skills**
4. Occupational reallocation = important margin of adjustment *[e.g. Dauth et al., 2021]*
   ⇒ **occupational choice**

Big picture: requirements → key model features

Q: Who will gain and who will lose from AI-induced task automation?

1. Technological change as task automation *[Acemoglu-Autor, 2011; Acemoglu-Restrepo, 2018; ...]*
   ⇒ **task-based framework**
2. Exposure to automation varies across tasks and hence occupations *[e.g. Webb, 2020]*
   ⇒ explicit model of **occupation-specific production technology**
3. Earnings effects from automation vary depending on skill portfolio *[e.g. Autor, 2015]*
   ⇒ **high-dimensional, task-specific skills**
4. Occupational reallocation = important margin of adjustment *[e.g. Dauth et al., 2021]*
   ⇒ **occupational choice**
5. ***Not today, in progress:*** GE/sectoral price effects

▸ Discussion

Environment: task-based production meets Roy

- Discrete time (t), repeated static model
- **Production technology:** task-based with occupation-specific weights
  - production is Cobb-Douglas over discrete task set $\mathcal{T}$
  - occupation $o \in \mathcal{O}$ **bundles tasks** with weights $\{\alpha_{o,\tau}\}_{\tau \in \mathcal{T}}$

Environment: task-based production meets Roy

- Discrete time (t), repeated static model
- **Production technology:** task-based with occupation-specific weights
  - production is Cobb-Douglas over discrete task set $\mathcal{T}$
  - occupation $o \in \mathcal{O}$ **bundles tasks** with weights $\{\alpha_{o,\tau}\}_{\tau \in \mathcal{T}}$
- **Firms:**
  - infinite supply of entrepreneurs who perfectly compete for a worker's labor
  - assign tasks ex-ante optimally to humans ($\rightarrow \mathcal{T}_l$) or machines w prod. $\{z_\tau\}_{\tau \in \mathcal{T}}$ ($\rightarrow \mathcal{T}_m$)
  - match with 1 worker, rent machines from inf. elastic capital market at exog. rate r

Environment: task-based production meets Roy

- Discrete time (t), repeated static model
- **Production technology:** task-based with occupation-specific weights
  - production is Cobb-Douglas over discrete task set $\mathcal{T}$
  - occupation $o \in \mathcal{O}$ **bundles tasks** with weights $\{\alpha_{o,\tau}\}_{\tau \in \mathcal{T}}$
- **Firms:**
  - infinite supply of entrepreneurs who perfectly compete for a worker's labor
  - assign tasks ex-ante optimally to humans ($\rightarrow \mathcal{T}_l$) or machines w prod. $\{z_\tau\}_{\tau \in \mathcal{T}}$ ($\rightarrow \mathcal{T}_m$)
  - match with 1 worker, rent machines from inf. elastic capital market at exog. rate r
- **Workers:**
  - log utility over consumption
  - heterogeneous, fixed **task-specific skills** $s_i = \{s_{i,\tau}\}_{\tau \in \mathcal{T}_l}$ where $\underset{\downarrow}{s_i} \sim \mathcal{N}(\bar{s}, \Sigma_s)$

    $|\mathcal{T}_l| \times 1$ *vector*
  - period t: draw shocks, choose occupation o, match with entrepreneur, produce & earn

Firm's optimal production problem

- **Output** of firm in occ o with worker i given idiosyncratic shock $\epsilon_{i,o,t} \sim \mathcal{N}(0, \varrho)$:

$$y_{i,o,t}(\cdot) = \underbrace{\prod_{\tau \in \mathcal{T}_l} \left(\exp\left(s_{i,\tau} + \varepsilon_{i,o,t}\right) \cdot \ell_{i,\tau,t}\right)^{\alpha_{o,\tau}}}_{\text{worker-produced}} \underbrace{\prod_{\tau \in \mathcal{T}_m} \left(\exp\left(z_\tau\right) \cdot m_{i,\tau,t}\right)^{\alpha_{o,\tau}}}_{\text{machine-produced}}$$

- **Profits:**

$$\pi_{i,o,t} = \max_{\{m_{i,\tau}\}_{\tau \in \mathcal{T}_m}, \{\ell_{i,\tau}\}_{\tau \in \mathcal{T}_l}} y_{i,o,t}\left(\{\ell_{i,\tau,t}\}_{\tau \in \mathcal{T}_l}, \{m_{i,\tau,t}\}_{\tau \in \mathcal{T}_m}\right) - \exp\left(w_{i,o,t}\right) - r \sum_{\tau \in \mathcal{T}_m} m_{i,\tau,t}$$

$$\text{s.t.} \sum_{\tau \in \mathcal{T}_l} \ell_{i,\tau,t} = 1$$

Firm's optimal production problem

- **Output** of firm in occ o with worker i given idiosyncratic shock $\epsilon_{i,o,t} \sim \mathcal{N}(0, \varrho)$:

$$y_{i,o,t}(\cdot) = \underbrace{\prod_{\tau \in \mathcal{T}_l} \left(\exp\left(s_{i,\tau} + \varepsilon_{i,o,t}\right) \cdot \ell_{i,\tau,t}\right)^{\alpha_{o,\tau}}}_{\text{worker-produced}} \underbrace{\prod_{\tau \in \mathcal{T}_m} \left(\exp\left(z_\tau\right) \cdot m_{i,\tau,t}\right)^{\alpha_{o,\tau}}}_{\text{machine-produced}}$$

- **Profits:**

$$\pi_{i,o,t} = \max_{\{m_{i,\tau}\}_{\tau \in \mathcal{T}_m}, \{\ell_{i,\tau}\}_{\tau \in \mathcal{T}_l}} y_{i,o,t}\left(\{\ell_{i,\tau,t}\}_{\tau \in \mathcal{T}_l}, \{m_{i,\tau,t}\}_{\tau \in \mathcal{T}_m}\right) - \exp\left(w_{i,o,t}\right) - r \sum_{\tau \in \mathcal{T}_m} m_{i,\tau,t}$$

$$\text{s.t.} \sum_{\tau \in \mathcal{T}_l} \ell_{i,\tau,t} = 1$$

- **Optimality:**

▸ FOC capital

$$\ell_{i,\tau,t} = \frac{\alpha_{o,\tau}}{\sum_{\tau \in \mathcal{T}_l} \alpha_{o,\tau}}$$

Firm's optimal production problem

- **Output** of firm in occ o with worker i given idiosyncratic shock $\epsilon_{i,o,t} \sim \mathcal{N}(0, \varrho)$:

$$y_{i,o,t}(\cdot) = \underbrace{\prod_{\tau \in \mathcal{T}_l} \left(\exp\left(s_{i,\tau} + \varepsilon_{i,o,t}\right) \cdot \ell_{i,\tau,t}\right)^{\alpha_{o,\tau}}}_{\text{worker-produced}} \underbrace{\prod_{\tau \in \mathcal{T}_m} \left(\exp\left(z_\tau\right) \cdot m_{i,\tau,t}\right)^{\alpha_{o,\tau}}}_{\text{machine-produced}}$$

- **Profits:**

$$\pi_{i,o,t} = \max_{\{m_{i,\tau}\}_{\tau \in \mathcal{T}_m}, \{\ell_{i,\tau}\}_{\tau \in \mathcal{T}_l}} y_{i,o,t}\left(\{\ell_{i,\tau,t}\}_{\tau \in \mathcal{T}_l}, \{m_{i,\tau,t}\}_{\tau \in \mathcal{T}_m}\right) - \exp\left(w_{i,o,t}\right) - r \sum_{\tau \in \mathcal{T}_m} m_{i,\tau,t}$$

$$\text{s.t.} \sum_{\tau \in \mathcal{T}_l} \ell_{i,\tau,t} = 1$$

- **Optimality:**

▸ FOC capital

$$\ell_{i,\tau,t} = \frac{\alpha_{o,\tau}}{\sum_{\tau \in \mathcal{T}_l} \alpha_{o,\tau}} \longrightarrow \textit{matrix A: } |\mathcal{O}| \times |\mathcal{T}_l|$$

Wage equation

Intercept term

$$w_{i,o,t} = \overbrace{\mu_o}^{\text{occ.-specific intercept}} + \overbrace{\sum_{\mathcal{T}_l} \frac{\alpha_{o,\tau}}{LS_o} \cdot s_{i,\tau}}^{\text{weighted skills}} + \overbrace{\varepsilon_{i,o,t}}^{\text{idiosyncratic productivity shock}}$$

$$= \mu_o + \underbrace{\frac{1}{n_{\text{skill}}} \sum_{\mathcal{T}_l} s_{i,\tau}}_{\text{scalar absolute advantage}} + \text{Cov}\left(n_{\text{skill}} \cdot \frac{\alpha_{o,\cdot}}{LS_o}, \underbrace{s_{i,\cdot} - \frac{1}{n_{\text{skill}}} \sum_{\mathcal{T}_l} s_{i,\tau}}_{\text{specialization vector}} \right) + \varepsilon_{i,o,t}$$

$\text{LS}_o = \sum_{\tau \in \mathcal{T}_l} \alpha_{o,\tau}$: labor share in occupation o

Occupational choice

- Each period, worker i chooses occ. subject to preference shock $u_{i,o,t} \sim \text{Gumbel}(0, \nu)$:

$$\hat{o}_{i,t} = \text{argmax}_o w_{i,o,t} + u_{i,o,t}$$

- **Occupational choice probabilities:**

$$P(\hat{o} = o | w_{i,\cdot,t}) = \frac{\exp(w_{i,o,t}/\nu)}{\sum_{o'} \exp(w_{i,o',t}/\nu)}$$

- No exogenous switching costs (currently)

Automation in the model

- **Automation** of task τ^*: a one-time, permanent rise in machine productivity $z_{\tau^\star}$ that is *just* large enough to make it optimal to reassign τ^* from humans to machines

$$\mathcal{T}_l' = \mathcal{T}_l \backslash \tau^\star \qquad \mathcal{T}_m' = \mathcal{T}_m \cup \tau^\star$$

  - NB: lower bound on positive productivity effects

Automation in the model

- **Automation** of task τ^*: a one-time, permanent rise in machine productivity $z_{\tau^\star}$ that is *just* large enough to make it optimal to reassign τ^* from humans to machines

$$\mathcal{T}_l' = \mathcal{T}_l \backslash \tau^\star \qquad \mathcal{T}_m' = \mathcal{T}_m \cup \tau^\star$$

  - NB: lower bound on positive productivity effects

- **Change** in expected log (potential) wage for i in occupation o

$$w_{i,o,t+1} - w_{i,o,t} = \mu_{o,t+1} - \mu_{o,t} + \frac{\alpha_{o,\tau^\star}}{LS_o}\left(\sum_{\mathcal{T}_l \backslash \tau^\star} \frac{\alpha_{o,\tau}}{LS_o - \alpha_{o,\tau^\star}} s_{i,\tau} - s_{i,\tau^\star}\right)$$

$\Rightarrow$ **A worker is more likely to win if *relatively* skilled in non-automated tasks**

Measurement

Theory meets data: overview

▸ Why not O*NET GWAs/importance weights

- **Goal:** parametrize the model at same 'resolution' as task exposure measures
- **Step 1:** map model tasks & occupations to data
  - use O*NET's $\sim$ 19,000 task statements ($\sim$ most exposure measures) & *cluster* them
  - occupations: 90+ SOC-2000 minor groups ($\sim$ 3d)
- **Step 2:** measure task weights for all occupations
  - task-weight matrix A with (o, τ) entry $\frac{\alpha_{o,\tau}}{\sum_{\tau \in \mathcal{T}_l} \alpha_{o,\tau}}$
- **Step 3:** estimate unobserved $(\bar{s}, \Sigma_s)$ using MLE
  - given A and NLSY '79

Steps 1 + 2: constructing the task-weight matrix *A*

Validation | Examples: occ | Examples: tasks

Composing Technical Documentat...
Evaluating and Strategizing
Performing Clinical and Labora...
Repairing and Maintaining Equi...
Performing Precision Fabricati...
Managing Food Service Operatio...
Managing Customer Communicatio...
Developing Technical Systems
Designing and Analyzing System...
Manipulating and Positioning M...
Supervisors, Office and Admini...
Financial Specialists
Health Diagnosing and Treating...
Extraction Workers
Assemblers and Fabricators
Supervisors, Food Preparation ...
Other Teachers and Instructors
Sales Representatives, Wholesa...
Drafters, Engineering, and Map...
Engineers
Other Healthcare Practitioners...
Nursing, Psychiatric, and Home...
Food and Beverage Serving Work...
Retail Sales Workers
Entertainment Attendants and R...
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

Step 3: estimation of task-specific skills

Details

- **Challenge:** skill distribution is unobserved
- Show how to estimate $(\bar{s}, \Sigma_s)$
  - variation: realized wages & occupational choices
  - intuition: economist vs software engineer
- **Data:** NLSY 79 + A matrix
  - worker-level panel of occupational choices and wages
- **Formalization:** max. likelihood
- **Implementation:** Monte Carlo integration + auto-diff. + stochastic gradient descent

Validation: Monte-Carlo study

(a) Means

(b) Covariance

Estimated skill correlation matrix

Model properties: selection based on comparative advantage

Transition vs. specialization

- Workers tend to select into occupations which load heavily on tasks they are relatively skilled at – example of *Top Executives*

relative specialization:

$$s_{i,\cdot} - \frac{1}{n_{\text{skill}}} \sum_{\mathcal{T}_l} s_{i,\tau}$$

Model properties: occupational transition probabilities

Wages and emp. shares

- Model endogenously generates (some) persistence and directionally tracks switching patterns

Model properties: occupational transitions reflect task requirements

- Workers are more likely to move to occupations with similar task requirements

[cf. Gathmann-Schoenberg, 2010]

Application

Webb's (2020) exposure measures

▸ Patent criteria ▸ Eloundou et al. (2023)

Historical shock: industrial robots

- **Literature:** ambiguous estimates of relationship b/w robot exposure and wages

  [e.g. Acemoglu-Restrepo, 2020; Aghion et al., 2020; Dauth et al., 2021]

Historical shock: industrial robots

- **Model:** ambiguous estimates of relationship b/w robot exposure and wages

AI: occupation-level effects of automating “analyzing complex data”

Yet for incumbents...

Mechanism: relative skill specialization shapes wage effects...

$s_{i,\tau^*} - \frac{1}{n_{\text{skill}}} \sum_{\tau \in \mathcal{T}_l} s_{i,\tau}$

...so selection yields negative slope for incumbents

⇒ As workers *select* into occupations by comparative advantage, high *occupational* exposure also *tends* to imply *relative skill specialization* in the automated task

Incumbents: stayers do better than switchers

- Consistent with evidence on task 'upgrading' for stayers *[Bartel et al., 2007; Dauth et al., 2021]* and losses for occupation switchers *[e.g. Huckfeldt, 2022]*

(a) Incumbents: stayers

(b) Incumbents: switchers

So why the positive effect at the occupational level?

- Consistent with evidence on positive wage effects from in-switching *[Artuc et al., 2010; Humlum, 2021]* but in model likely overstated (no GE) & too fast (no switching frictions)

Summing up: AI

Conclusion

Summary

- **Early stage – feedback very very welcome!**

- **Core contribution:** empirically rich tractable framework to quantify (and forecast) who wins and who loses from (AI induced) task automation
  - forward-looking element critical for policy

- **Key insight:** automation harms you if you are specialized in the automated task – *but* benefits you if freed up to pursue tasks in which you're more productive

  $\Rightarrow$ **automation effects depend on *which* task is being automated**

- **Next steps:**
  - model: GE
  - more counterfactuals: self-driving vehicles
  - AI regulation: minimum human-in-the-loop requirements

Discussion

Extra Slides

Automated document review: good or bad?

Back

GE: plan

Back

- Missing important model feature: heterogeneous, endogenous occupation prices
  - steady-state: high-wage occ's involve scarce skills hence high o price
  - counterfactual: occupational price response as a function of demand elasticities
- Identification challenge: μ_o becomes endogenous and the following equation is satisfied by more than one pair $(\mu_o, \bar{s})$:

$$\text{mean potential wage}_o = \mu_o + A'_{o,\cdot} \circ \bar{s}$$

  where $\bar{s}$ is vector of average skills
- Options we're exploring:
  1. time variation in task shares
  2. dynamic skill accumulation
  3. identifying restriction $A \perp \mu_o$

FOCs

Back

- FOC for machines $m := \sum_{\tau \in \mathcal{T}_m} m_\tau$:

$$\left(\sum_{\tau \in \mathcal{T}_m} \alpha_{o,\tau} \right) \frac{y}{r} = m$$

and

$$m_\tau = \frac{\alpha_{o,\tau}}{\sum_{\tau \in \mathcal{T}_m} \alpha_{o,\tau}} m$$

- Given

$$\log y_o = \left[\sum_{\tau \in \mathcal{T}_l} \frac{\alpha_{o,\tau}}{\sum_{\tau \in \mathcal{T}_l} \alpha_{o,\tau}} s_{i,\tau} \right] + \varepsilon_{i,o}$$

$$+ \left[\sum_{\tau \in \mathcal{T}} \frac{\alpha_{o,\tau}}{\sum_{\tau \in \mathcal{T}_l} \alpha_{o,\tau}} \log(\alpha_{o,\tau}) \right] - \log \left(\sum_{\tau \in \mathcal{T}_l} \alpha_{o,\tau} \right) + \left[\sum_{\tau \in \mathcal{T}_m} \frac{\alpha_{o,\tau}}{\sum_{\tau \in \mathcal{T}_l} \alpha_{o,\tau}} (z_\tau - \log r) \right],$$

Wage equation: details

Back

- Intercept

$$\mu_o = \sum_{\tau \in \mathcal{T}} \frac{\alpha_{o,\tau}}{\sum_{\tau \in \mathcal{T}_l} \alpha_{o,\tau}} \log\left(\alpha_{o,\tau}\right) + \left(\sum_{\tau \in \mathcal{T}_m} \frac{\alpha_{o,\tau}}{\sum_{\tau \in \mathcal{T}_l} \alpha_{o,\tau}} \left(z_\tau - \log r\right)\right)$$

- We assume that in the initial steady state there is only one composite machine task with productivity normalized to $\log r$, which implies that μ_o is known for all occupations.

Details on the estimation strategy I

Back

- Exact likelihood:

$$\prod_i \int_s \left[\left(\int_{w_{i,\cdot,-\omega_\cdot}} \prod_t P(\hat{o}_{i,t} = \omega_{i,t} | w_{i,\cdot,\cdot}, \nu) \cdot f(w_{i,t,-\omega_t} | s, w_{i,\cdot,\omega_\cdot}, \varsigma) \right) \cdot f(s | w_{i,\cdot,\omega_\cdot}, \varsigma, \bar{s}, \Sigma_s) \right]$$
$$\cdot f(w_{i,\cdot,\omega_\cdot} | \varsigma, \bar{s}, \Sigma_s)$$

- **Strategy:** Monte Carlo integration - for all i generate n_o draws from

$$f(w_{i,\cdot,-\omega_\cdot} | w_{i,\cdot,\omega_\cdot}, \varsigma, \bar{s}, \Sigma_s) = \int_s f(w_{i,\cdot,-\omega_\cdot} | s, w_{i,\cdot,\omega_\cdot}, \varsigma) f(s | w_{i,\cdot,\omega_\cdot}, \varsigma, \bar{s}, \Sigma_s)$$

and evaluate the mean of $P(\hat{o}_{i,t} = \omega_{i,t} | w_{i,\cdot,t}, \nu)$ to obtain an estimator for $\mathcal{L}_i(\theta)$:

$$\hat{\mathcal{L}}_i(w_{i,t,\omega}, \nu, \varsigma, \bar{s}, \Sigma_s) = \left(\frac{1}{n_o} \sum_j \prod_t P(\hat{o}_{i,t} = \omega_{i,t} | w_{j,t,\cdot}, \nu) \right) \cdot f(w_{i,\cdot,\omega_\cdot} | \varsigma, \bar{s}, \Sigma_s)$$

Details on the estimation strategy II

Back

- Two numerical techniques help speed up the maximum likelihood computation
- **Auto-differentiation:** efficiently compute the gradient of this function
- **Stochastic gradient descent:**
  - basic technique: gradient descent

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla \left(-\mathcal{L}(\theta_t)\right)$$

  - randomly partition individuals into n groups:

$$\{1, 2, \ldots, I\} = B_1 \cup B_2 \cup \ldots \cup B_n, \quad B_i \cap B_j = \emptyset$$

  - calculate the likelihood based on batch $B_1, \ldots, B_n$ only
  - when done, draw a new partition

Why not use O*NET GWAs and importance weights

Back

- Potential alternative to our approach: use O*NET "General Work Activities" (GWAs) and occupational importance weights
- Reasons we prefer our approach:
  1. GWAs themselves are not mutually exclusive (e.g. "Analyzing Data or Information" vs "Processing Information") nor exhaustive (esp. regarding activities differentiating high-wage occupations, e.g. complex quantitative analyses), and some seem ambiguous ("Getting Information")
  2. Weights available (importance/level/frequency) don't correspond to time shares, as required to map onto the theory
  3. GWAs + LLM-generated time shares: resulting A matrix is low-rank ($\rightarrow$ poor model fit)
  4. Flexibility: our approach is consistent with different occupational classifications (e.g. SOC-2000, which can be x-walked to NLSY) and time periods

Validation of LLM-generated time shares: overview

1. Comparison of time share measurement: LLM vs BIBB survey ✓
2. Comparison of LLM-generated time shares for GWAs to O*NET importance weights ✓
3. Internal consistency: do measurements for detailed occupations aggregate up? ✓

- What else would you like us to check?
  – comparison across LLMs?

Validation: LLM-generated task shares vs. BIBB

(a) Occupation-level correlations

(b) Task-level correlations

Validation: O*NET GWAs (1)

- Take O*NET GWAs (O*NET 5.0, consistent with SOC-2000), construct relative importance for each GWA by occupation, aggregate to SOC-2000-3d
- Let LLM generate *time shares* for the GWAs for each SOC-2000-3d occ
- How do LLM-time shares correlate with vector of O*NET importance weights?

Validation: O*NET GWAs (2): correlation across occupations by task

Validation: internal consistency

Model properties: transition probabilities decline in specialization

Back

- Workers with v specialized (= dispersed) skills are less likely to switch occupation

Model fit: occupational wages and employment shares

Back

A matrix: example tasks - extracted skills - tasks

Back

Task	Activity	Skills	Cluster
Direct or coordinate an organization's financial or budget activities to fund operations, maximize investments, or increase efficiency	Direct financial operations	Financial management (expert), strategic planning (advanced), budgeting (advanced), analytical thinking (advanced)	Evaluating and Strategizing
Clean and sterilize vats and factory processing areas	Clean and sterilize processing areas	Manual dexterity (basic)	Performing Material Handling Tasks
Press switches and turn knobs to start, adjust, and regulate equipment, such as beaters, extruders, discharge pipes, and salt pumps	Operate equipment controls	Technical knowledge (intermediate), manual dexterity (basic)	Performing Precision Technical Tasks
Conduct research, data analysis, systems design, or support for software such as Geographic Information Systems (GIS) or Global Positioning Systems (GPS) mapping software	Conduct research and data analysis for GIS software	Research skills (advanced), data analysis (advanced), systems design (advanced)	Analyzing Complex Data

A matrix: example occupations

Back

- Evaluating and Analyzing Information
- Documenting and Organizing Information
- Designing and Delivering Instruction
- All other tasks

- Analyzing Financial and Business Data
- Evaluating and Analyzing Information
- Analyzing Complex Data
- All other tasks

Webb measure: selection criteria

Back

Table A1: **Patent selection criteria.**

Technology	Definition
AI	Title/abstract include "neural network", "deep learning", "reinforcement learning", "supervised learning", "unsupervised learning", or "generative model"
Software	Title/abstract include "software", "computer", or "program" AND title/abstract exclude "chip", "semiconductor", "bus", "circuit", or "circuitry"
Robots	Title/abstract include "robot"

Notes: Patents corresponding to each technology are selected using these keyword inclusion/exclusion criteria.

Aggregated task exposure measures from Eloundou et al. (2023)

Webb (2020)

Webb's historical evidence on effects of robots

Back

The ins and outs of occupations: robots

Back

(a) Incumbents

(b) In-switchers

LLMs: occupation-level effects (2) - alternative task

The role of relative skill specialization in the automated task

- Reallocation across tasks *[Bartel et al., 2007; Dauth et al., 2021]*