On the Identification of Models of Uncertainty, Learning, and Human Capital Acquisition and the Determinants of Sorting

Aureo De Paula [1] Cristina Gualdani[2] Elena Pastorino[3] Sergio Salgado[4]

[1]UCL [2]Queen Mary [3]Stanford and Hoover [4]Wharton

November 2023

Introduction

- Question: How come measured degree of sorting so low if canonical models of inequality predicated on it?
    - ▷ Examples: estimates in AKM (1999), CHK (2013) and similar

- Answer: Models of sorting with human capital acquisition and (symmetric) learning about productivity
    - ▷ Naturally give rise to countervailing (compensating differential) forces
    - ▷ Option value of HK and info acquisition attenuates impact of firm/worker/match effects on wages
    - ▷ Idea: "Rosen meets sorting"

- Examples
    - ▷ Bad jobs (low firm effect) can pay high wages and attract productive workers
    - ▷ Good jobs at which a lot of HK/information can be accumulated can pay low wages

- Critical to *understand* the determinants of these models is being able to *recover* them
    - ▷ Difficult: they are complex unobserved time-varying processes endogenous functions of primitives

Focus of the Analysis

- Consider general class of equilibrium models with learning, HK acquisition and rich heterogeneity
  - ▷ Account for imperfect competition (Bertrand) among differentiated firms (robust to firm entry)
  - ▷ Capture situations in which potential outcomes of interest are unobserved
  - ▷ Allow for dynamic selection based on observables and unobservables

- Show they are identified from wage and job information (builds on Pastorino (2022))

- Revisit determinants of sorting

Data and Primitives of Interest

- **Data**: for T periods and a large sample of i.i.d. agents, the econometrician
  - ▷ Observes wage, employing firm/job and initial human capital
  - ▷ Does **not** observe skill type, ability, beliefs about ability, **output (signal)** or human capital investments

- **Primitives**: to be identified in order to perform any counterfactual of interest
  - ▷ Skill type distribution
  - ▷ Prior and signal distribution (learning process)
  - ▷ Conditional firm/job assignment probabilities (CCPs)
  - ▷ Expected wage, output and human capital process
  - ▷ Discount factor, expected wage, output and human capital parameters

Key Challenges

- Dynamic equilibrium model of *selection* on observables and unobservables with two-sided heterogeneity
- Outcome equations (equilibrium wages and CCPs) are *nonlinear* in parameters and factors
- Agents learn from outcomes (signals) *unobserved* to the econometrician
- CCPs
  - ▷ Are conditional on these unobserved signals
  - ▷ The impact of signals on CCPs is mediated by wages in a nonlinear way
- Agents' prior beliefs may be *subjective* i.e. may not coincide with true distribution of latent factors

Key Identification Results

- Relax linearity of outcome equations

- In addition to component densities, identify CCPs from wage distribution alone
  - ▷ From weights of wage mixture distribution over firm/job/signal histories, skill types and initial HK
  - ▷ Outcomes, signals, types and priors may be discrete or continuous

- From recovered CCPs and wage mixture densities, identify all other primitives

- Importantly: identify prior and signal distributions w/o parametric restrictions on learning process

Roadmap of Identification Argument

- Identify **wage mixtures** over all possible firm/job/signal histories, skill types and initial HK

- From wage **mixture weights**, identify **skill type**, **initial prior** and **signal** distributions as well as **CCPs**
  - ▷ From signal distribution, identify **belief** process
  - ▷ From initial prior and signal distributions, the **learning** process is pinned down

- From wage **mixture densities**, identify **wage**, **output** and **HK processes**
  - ▷ Current version: output linear in HK and law of motion of HK unaffected by signals (can relax)

- From **structural model**, identify **discount factor**, **wage**, **output** and **HK parameters**

A Review of the Wage Equation

- Dynamic equilibrium model of labor market w/ N firms competing for workers over horizon of length $T \leq \infty$
  - ▷ Through simultaneous offer of jobs (occupations) and wages each period
  - ▷ We focus on Markov perfect equilibria

- At time t the wage paid to a worker is

$$w_t = y(s_{i,t}, g, k_g) + \Psi(s_{i,t}, f, k_f, g, k_g) + \epsilon_{i,g,k_g,t}$$

  - ▷ f is **employing** firm at job k_f, g is **second-best** firm (offering second-highest PV of wages) at job k_g
  - ▷ $s_{i,t} := (i, p_t, \kappa_t)$ is deterministic state
    - ∗ i is **skill type**
    - ∗ p_t is **belief** that worker has high ability at beginning of t (high/low ability for expositional simplicity)
    - ∗ κ_t is **initial human capital** (h_1) and human capital **investments** up to $t-1$
  - ▷ $y(s_{i,t}, g, k_g)$ is **output** at job k_g of firm g and $\Psi(s_{i,t}, f, k_f, g, k_g)$ is **compensating differential**
  - ▷ $\epsilon_{i,g,k_g,t}$ captures **productivity shocks** and **measurement errors**

Wage Equation: Components and Interpretation

- Sum of three terms
  - ▷ $y(s_{i,t}, g, k_g)$: affine in $s_{i,t}$ (interactive fixed effect extension of AKM)
  - ▷ $\Psi(s_{i,t}, f, k_f, g, k_g)$: compensating differential that dampens the impact of fixed effects on wages
  - ▷ $\epsilon_{i,g,k_g,t}$: idiosyncratic factors (**can use equilibrium to express k_g as fct of k_f**)

- The second term explains why sorting is typically estimated to be *low*
  - ▷ Captures expected PV difference in match surplus from working at employing firm vs. second-best firm
  - ▷ This term is higher the higher the value of HK or info worker forgoes by choosing to work at f
  - ▷ So *can compensate* for small $y(s_{i,t}, g, k_g)$

Useful Notation for Next Arguments

- At the end of time t worker's signal is high or low: $R_t \in \{0,1\}$ and $R^t := (R_1, R_2, \ldots, R_t)$
  - ▷ This *performance* signal is binary for simplicity, continuous signals can be **easily handled**
  - ▷ R_t depends on skill type, ability, firm/job histories and human capital

- $L_{f,k_f,t}$ takes value 1 if f is employing firm at job k_f and time t; zero otherwise

- $L_{g,k_g,t}$ takes value 1 if g is second-best firm at job k_g and time t; zero otherwise

- $L_t := (L_{f,k_f,t} : f \in \mathcal{F}, k_f \in \mathcal{K}_f)$, $L_{t,2} := (L_{g,k_g,t} : g \in \mathcal{F}, k_g \in \mathcal{K}_g)$ and $L^t := (L_1, L_2, \ldots, L_t)$

Identification of Wage Mixture

- According to the model, $(s_{i,t}, L_{t,2})$ are functions of (L^t, h_1, i, R^{t-1}) so the wage mixture is

$$g(w_t|L^t, h_1) = \sum_{R^{t-1}\in\{0,1\}^{t-1}} \int_{i\in\mathcal{I}} g(w_t|L^t, h_1, i, R^{t-1}) q(i, R^{t-1}|L^t, h_1) di$$

- Assume
  - ▷ $\epsilon_{i,L_{t,2},t}$ are i.i.d. with distribution conditional on (L^t, h_1, i, R^{t-1}) denoted by $r(\cdot|L^t, h_1, i, R^{t-1})$
  - ▷ $\mathcal{I}$ is finite with known cardinality
  - ▷ $r(\cdot|L^t, h_1, i, R^{t-1})$ is mixture of Normals with unknown means/variances in compact set $D(L^t, h_1, i, R^{t-1})$

- By Bruni and Koch (1985): $\{g(\cdot|L^t, h_1, i, R^{t-1}), q(i, R^{t-1}|L^t, h_1)\}$ are identified up to labelling wrto (i, R^{t-1})

- Can accommodate continuous signals and/or skill types (no parametric restriction on components) details

Exchangeability of Mixture Components

- Pairs $\{g(\cdot|L^t, h_1, i, R^{t-1}), q(i, R^{t-1}|L^t, h_1)\}$ are equivalent across R^{t-1} with same no. of successes by job

- To handle such *exchangeability*
  - ▷ Group realized R^{t-1} into equivalence classes by job/success count (other suff. stat. in continuous case)
  - ▷ Select one representative element from each equivalence class
  - ▷ Proceed with dimension reduction of wage mixture before delving into its identification

Labelling Mixture Components

- Bruni and Koch (1985) allow us to identify pairs $\{g(\cdot|L^t, h_1, i, R^{t-1}), q(i, R^{t-1}|L^t, h_1)\}$

- However, we are unable to label each pair with respect to (i, R^{t-1})

- Labelling is crucial for identifying CCPs and learning process
  - ▷ These objects are identified from mixture weights by applying law of total probability
  - ▷ It requires us to "integrate over" the correct components (more on this later)

- We solve this issue by ordering component densities according to their means and/or variances

Identification of Skill Type Distribution

- Direct implication of the wage mixture identification

- Consider the wage mixture at $t = 1$

$$g(w_1|L_1, h_1) = \int_{i \in \mathcal{I}} g(w_1|L_1, h_1, i) q(i|L_1, h_1) di$$

- From the mixture weights, we identify the skill type distribution $q(\cdot|L_1, h_1)$

Identification of Signal Distribution

- Usually the signal distribution is unobserved to the econometrician: we maintain so

- Key idea: recover this distribution (also CCPs as shown next) from wage mixture weights

- Simple algebra gives

$$\begin{aligned}\Pr(R^{t-1}|L^{t-1},h_1,i) &= \sum_{L_t\in\mathcal{L}_t}\Pr(L_t|L^{t-1},h_1,i,R^{t-1})\Pr(R^{t-1}|L^{t-1},h_1,i)\\ &= \sum_{L_t\in\mathcal{L}_t}\frac{\Pr(L^t,h_1,i,R^{t-1})}{\Pr(i|L^{t-1},h_1)\Pr(L^{t-1},h_1)}\end{aligned}$$

- Objects in last expression are identified from mixture weights ($\Pr(i|L^{t-1},h_1)$ from weights in $t>1$) and data

- Thus, the performance distribution $\Pr(R^{t-1}|L^{t-1},h_1,i)$ is identified

Identification of CCPs

- $\Pr(L_1|h_1, i)$ is directly identified from mixture weights in period 1

- In period $t > 2$ simple algebra gives

$$\Pr(L_t|L^{t-1}, h_1, i, R^{t-1}) = \frac{\Pr(L^t, h_1, i, R^{t-1})}{\Pr(R^{t-1}|L^{t-1}, h_1, i)\Pr(i|L^{t-1}, h_1)\Pr(L^{t-1}, h_1)}$$

- Objects on RHS are identified from mixture weights, signal distribution and data

- Thus, $\Pr(L_t|L^{t-1}, h_1, i, R^{t-1})$ is identified

Identification of Learning Process

- Workers can have either high or low ability $\theta \in \{\bar{\theta}, \underline{\theta}\}$ (straightforward to extend to continuous case)

- For $t > 2$ apply Bayes rule to get beliefs $\{p_t\}$ with $p_t := \Pr(\theta = \bar{\theta} | L_{f,k_f,t} = 1, L^{t-2}, h_1, i, R^{t-1})$ or

$$p_t = \left[\frac{\alpha_{f,k_f} p_{t-1}}{\alpha_{f,k_f} p_{t-1} + \beta_{k,f}(1 - p_{t-1})}\right]^{R_{t-1}} \left[\frac{(1 - \alpha_{f,k_f}) p_{t-1}}{(1 - \alpha_{f,k_f}) p_{t-1} + (1 - \beta_{k,f})(1 - p_{t-1})}\right]^{1 - R_{t-1}}$$

- Key parameters governing learning process
  - ▷ Prior $p_1(f, k_f, \bar{h}, \iota) := \Pr(\theta = \bar{\theta} | L_{f,k_f,1} = 1, h_1 = \bar{h}, i = \iota)$
  - ▷ $\alpha_{f,k_f} := \Pr(R_t = 1 | \theta = \bar{\theta}, L_{f,k_f,t} = 1)$
  - ▷ $\beta_{f,k_f} := \Pr(R_t = 1 | \theta = \underline{\theta}, L_{f,k_f,t} = 1)$

Identification of Learning Process

- More on discrete example: $\Pr(R^t | L_{f,k_f,1} = \cdots = L_{f,k_f,t} = 1, h_1 = \bar{h}, i = \iota)$ is 2-component Binomial mixture

$$\Pr(R^t | L_{f,k_f,1} = \cdots = L_{f,k_f,t} = 1, h_1 = \bar{h}, i = \iota)$$
$$= \alpha_{f,k_f}^{\sum_{j=1}^t R_j} (1 - \alpha_{f,k_f})^{(t - \sum_{j=1}^t R_j)} p_t(f, k_f, \bar{h}, \iota) + \beta_{f,k_f}^{\sum_{j=1}^t R_j} (1 - \beta_{f,k_f})^{(t - \sum_{j=1}^t R_j)} [1 - p_t(f, k_f, \bar{h}, \iota)]$$

- LHS is identified from the identification of the performance distribution

- By Blischke (1964; 1978): mixture weights and probabilities are identified if # experiments ≥ 3

- Thus, we identify $(\alpha_{f,k_f}, \beta_{f,k_f})$ from $\Pr(R^3 | L_{f,k_f,1} = L_{f,k_f,2} = L_{f,k_f,3} = 1, h_1 = \bar{h}, i = \iota)$

- We obtain initial prior from $\Pr(R_1 = 1 | L_{f,k_f,1} = 1, h_1 = \bar{h}, i = \iota) = \alpha_{f,k_f} p_1(f, k_f, \bar{h}, \iota) + \beta_{f,k_f} [1 - p_1(f, k_f, \bar{h}, \iota)]$

- Once signal distribution recovered, its parameters too provided signal distribution *identifiable mixture* of them

Identification of Wage, Output and HK Processes

- Expected wage $y(s_{i,t}, g, k_g) + \Psi(s_{i,t}, f, k_f, g, k_g)$ is identified from mixture densities if shocks mean zero

- Expected output $y(s_{i,t}, g, k_g)$ is identified from expected wages
  - ▷ Under appropriate location normalizations
  - ▷ By exploiting parametric form of $y(\cdot)$: it is an affine function of the state by construction

- HK process is identified from process of deterministic state and CCPs

Identification of Discount Factor

- Combine information on job choices (discrete controls) and wages (continuous controls)

- Rely on exchangeability of time-varying latent process (beliefs)

- Exchangeability implies the recursed-out component mean wage is eventually polynomial of low order in δ
  - ▷ Past some period, the HK process dies out
  - ▷ In short panels: past some period, HK is accumulated symmetrically across occupations

Identification of Wage, Output and HK Parameters

- We can write our wage equation at time t for worker n as

$$w_{n,t} = \alpha_{n,k_g} + \psi_{g,k_g} + s_{n,t}^{\top}\gamma_{g,k_g} + \Psi(s_{n,t}, f, k_f, g, k_g) + \epsilon_{n,g,k_g,t}$$

- Compare with wage equation in AKM

$$w_{n,t} = \alpha_n + \varphi_{k_f} + x_{n,t}^{\top}\beta + \epsilon_{n,t}$$

  - ▷ Standard fixed effect formulation
  - ▷ Sorting measured by $\mathrm{Cov}(\alpha_n, \varphi_{k_f})$

Identification Strategy

- From identification perspective, three main differences wrto AKM wage equation
  - ▷ Equation nonlinear in unobserved effects in ways not captured by interactive formulation
  - ▷ Firm and worker fixed effects are job-specific
  - ▷ Second-best firm and state are unobserved

- So AKM identification proof does not straightforwardly apply

- We show identification of wage parameters by combining
  - ▷ AKM identification arguments
  - ▷ Identification results from structural model

- The identification of the output and HK parameters follows from $y(\cdot)$

Implications for Sorting

- Applying the usual variance decomposition arguments to our wage equation

$$\mathrm{Var}(w_{n,t}) = \underbrace{\mathrm{Cov}(\alpha_{n,k_g}, w_{n,t})}_{\text{person-job effect}} + \underbrace{\mathrm{Cov}(\psi_{g,k_g}, w_{n,t}) + \mathrm{Cov}(s_{n,t}^{\top}\gamma_{g,k_g}, w_{n,t})}_{\textit{static}\text{ firm-job effect}}$$

$$+ \underbrace{\mathrm{Cov}(\Psi(s_{n,t}, f, k_f, g, k_g), w_{n,t})}_{\textit{dynamic}\text{ firm-job effect (compensating differential)}} + \mathrm{Cov}(\epsilon_{n,g,k_g,t}, w_{n,t})$$

- We are investigating degree to which sorting occurs and extent to which is due to
  - ▷ Static vs. dynamic firm-job effects using U.S. employer-employee data (Census LEHD)
  - ▷ Key: common measures tend to over (under)-state sorting in short (long) run

- Next: examine impact of all these sources of sorting on dynamics of wage inequality in U.S.

Conclusion

- Examine general class of equilibrium models with learning, HK acquisition and rich heterogeneity
  - ▷ Account for imperfect competition among differentiated firms (robust to firm entry)
  - ▷ Capture situations in which potential outcomes of interest are unobserved
  - ▷ Allow for dynamic selection based on observables and unobservables
- Establish identification of latent processes, CCPs and primitive parameters
  - ▷ Based just on information on job choices and wages
- Help reconcile low estimates of sorting with sorting models for persistent wage inequality
  - ▷ What the literature has missed: if labor markets are to any extent competitive
  - ▷ Wages endogenously arbitrage differences in worker productivity across jobs and firms
  - ▷ So wages imperfectly measure sorting (only in $\simeq$ perfectly competitive models wages reflect productivity)
- Estimate degree of sorting and its evolution over time on Census LEHD data (in progress)

Identification of Wage Mixture with Continuous Types

- Suppose i is continuous
- Assume
  - ▷ $r(\cdot|L^t, h_1, i, R^{t-1})$ is Normal
  - ▷ The set of means and variances of $r(\cdot|L^t, h_1, i, R^{t-1})$ across i denoted by $D(L^t, h_1, R^{t-1})$ is compact
- Then, the wage mixture is

$$g(w_t|L^t, h_1) = \sum_{R^{t-1} \in \{0,1\}^{t-1}} \Pr(R^{t-1}|L^t, h_1) \int_{D(L^t, h_1, R^{t-1})} g(w_t|L^t, h_1, i, R^{t-1}; \mu, \sigma^2) d\pi(\mu, \sigma^2)$$

  - ▷ π is a probability measure over $D(L^t, h_1, R^{t-1})$
- By Bruni and Koch (1985): π and $\Pr(R^{t-1}|L^t, h_1)$ are identified back