

# Within Occupation Schooling Dispersion, Overeducation and Mismatch in the Labor Market: Theory and Empirics

JIE MA  
Colgate University



## Abstract

Concerns persist for years about whether individuals acquire more education than is required for their work, a phenomenon known as 'overeducation'. Ever since Duncan and Hoffman's seminal work (1981), previous literature documents mixed evidence and interprets it as evidence for inefficiency. To reconcile the contrasting facts, this paper builds a vertical schooling and occupation sorting model based on human capital, where education substitutes for ability. Both education and occupation choices are efficient in the theoretical model. I then use simulated data from the calibrated model to show that it reproduces patterns of estimates found in the literature. These estimates are in fact fully consistent with efficient decision making. Finally, I add lifecycle, information frictions and employer learning to the model to derive novel implications about the dynamics of education-job match. The paper then turns to the NLSY79 data to demonstrate that empirical evidence in the US from 1982-1994 is consistent with the theoretical model's predictions. Both the theoretical model predictions and the new empirical evidence rationalize the observed overeducation without implications of misallocation.

## Introduction: Overeducation Concerns

- Overeducation: individuals have more education than it is necessary for their occupations.
  - A worker with a bachelor's degree works as a Starbucks barista.
- Duncan and Hoffman's Specification
  - $\log(w_{it}) = \alpha^r S_{it}^r + \alpha^o S_{it}^o + \alpha^u S_{it}^u + X_{it} \beta + \epsilon_{it}$
  - $S_{it}^r$ : years of required school by occupation;  $S_{it}^u$ : years of underschooling ( $S_{it}^u = S_{it}^r - educ_{it}$ );  $S_{it}^o$ : years of overschooling  $S_{it}^o = educ_{it} - S_{it}^r$ .
  - Meta analysis of 151 studies (32 for North America, 94 for Europe, 18 for Asia) finds:  $\alpha^o = 4.3\% < \alpha^r = 8.9\%$
  - The ex-post heterogeneity in returns to different educational components is interpreted as evidence of inefficiency and misallocation.

## Research Questions

- Are the 'observed within occupation schooling dispersion' and lower returns to the 'surplus' schooling evidences of skill mismatch and inefficient allocation?
- Does the within occupation schooling difference imply that some individuals' education investment are sub-optimal? **No**
- What can be learned about a worker's human capital function?
  - Cognitive Ability v.s. Schooling: **Substitutes or Complements?**
- What can we learn about the labor market frictions, or other competing labor market theories?
  - Information Friction v.s. Search Friction? Human Capital v.s. Signaling?**

## Model: Static & Full Information

- A finite number of occupations, indexed by  $k \in \{0, 1, 2, \dots, K\}$ . Occupations are ranked in the order of increasing output prices.  $P_0 < P_1 < P_2 < \dots < P_{K-1} < P_K$
- Workers differ by human capital  $H_i = A_i + \alpha S_i$ . Workers sort into different occupations according to the single-dimension human capital index  $H_i$ .
- An occupation wants to obtain a fixed profit  $\Pi_k$  to any worker who is willing to take this contract.  $w_{ik}(H_i) = P_k H_i - \Pi_k$
- High ranked occupations extract higher profits.  $\Pi_0 < \Pi_1 < \Pi_2 < \dots < \Pi_{K-1} < \Pi_K$
- Occupation Sorting Rule: Worker  $H_i$  sort into occupation  $k$  if  $B_k < H_i \leq B_{k+1}$  where  $B_k = \frac{\Pi_k - \Pi_{k-1}}{P_k - P_{k-1}}$
- Cost of schooling is  $C(S_i, A_i) = \frac{cS_i^2}{2A_i}$

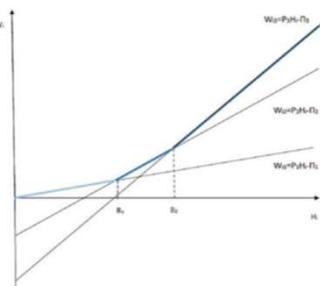


Figure 1. Illustration of Vertical Sorting

## Static Model Implications

**Proposition 1:** More able workers sort into higher ranked occupations.

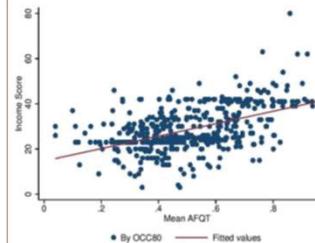


Figure 2. Occupation Sorting Based on Ability

**Proposition 2:** Conditional on occupation choice  $k$ , and labor market experience  $t$ , overeducated workers on average have lower ability.

$$\text{Prob}(y = \text{over}) = \beta_0 + \beta_1 \text{AFQT} + X\gamma + \epsilon$$

$$\beta_1 < 0$$

	All	>=16	>=18
AFQT	-0.263***	-0.465***	-1.150***
Rotter	-0.003	0.004	-0.004
Sociability	-0.005	0.020	0.005

Note. Model also includes controls for labor market experience, local unemployment rate, mother and father's education level in 1979, a dummy for born in US, a dummy for living in an urban area at the time of interview, a set of dummy variables for census regions, number of jobs had, weeks unemployed last year, gender, degree type. \*p<0.1. \*\*p<0.05. \*\*\*p<0.01

## Calibration & Simulation

**Test Assumption:** occupations pay heterogeneous return to human capital.

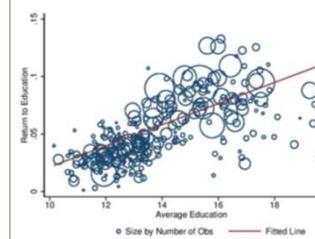


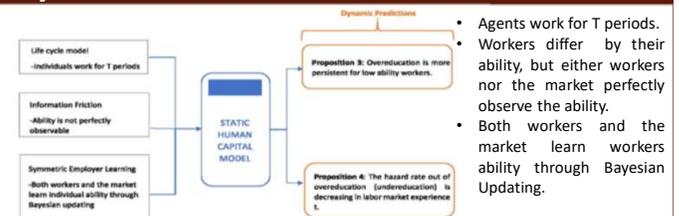
Figure 3. Evidence of Heterogeneous Return

**Simulation Result**

	$\alpha^r$	$\alpha^o$	$\alpha^u$
Simulated	0.078***	0.048***	-0.024***
Literature Range	[0.043, 0.135]	[-0.031, 0.054]	[-0.056, -0.025]
Mata (151 Studies)	0.089***	0.043***	-0.036***
Mata (US)	0.083***	0.046***	-0.027***

**Conclusion:** ex post heterogeneous returns may be due to model specification issues, can not be directly use as evidence against human capital model.

## Dynamic Model with Information Frictions



- Agents work for T periods.
- Workers differ by their ability, but either workers nor the market perfectly observe the ability.
- Both workers and the market learn workers ability through Bayesian Updating.

## Test Dynamic Model Predictions

**Proposition 3:** Overeducation is more persistent for low ability workers.

Define:  $P^+ = \text{Prob}(\mathbb{E}(H_{it+1}) > B_{k+1} | \mathcal{F}_t)$   
 $P^+$  is the probability that an overeducated individual  $i$  switches up to higher rank occupation  $\rightarrow$  out of overeducation.

$$\frac{\partial P^+}{\partial A_i} > 0$$

$$\text{Overeducation Spell} = \beta_0 + \beta_1 \text{AFQT} + X\gamma + \epsilon$$

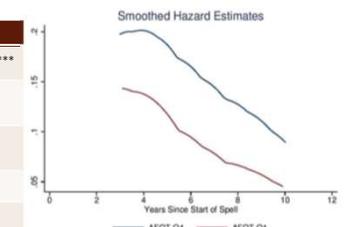
$$\beta_1 < 0$$

AFQT	-1.887***	-1.793***	-1.921***
Unemployment Rate	X	X	X
Rotter & Sociability		X	X
Experience		X	X
Demographic & Regional			X

**Proposition 4:** The hazard rate out of overeducation (undereducation) is decreasing in labor market experience  $t$ .

Define:  $P^+ = \text{Prob}(\mathbb{E}(H_{it+1}) > B_{k+1} | \mathcal{F}_t)$   
 $P^- = \text{Prob}(\mathbb{E}(H_{it+1}) \leq B_k | \mathcal{F}_t)$

$$\frac{\partial P^+}{\partial t} < 0 \text{ and } \frac{\partial P^-}{\partial t} < 0$$



## Conclusions

- Low ability workers acquire more schooling to compensate for their innate abilities  $\rightarrow$  Rationalize the observed within occupation schooling dispersion without the implication of sub optimal educational investment.
- Had those low ability workers not obtained enough schooling, they would end up in lower ranked occupations.  $\rightarrow$  Worse off.
- The conventional wage specification in the literature suffers from specification errors  $\rightarrow$  can not be interpreted as evidence for misallocation or against the human capital model.

### Contact

JIE MA Colgate University  
Email: jma@colgate.edu

Website: www.jiemacon.com