College Majors and Labour Market Mismatch

Michelle Rendall, Satoshi Tanaka, Yi Zhang

Monash University and CEPR, University of Queensland, University of Queensland

@AWPF, August 4, 2022

Motivation

- Universities offer various fields of study that students freely choose

- Hard to predict which skills will be demanded in labour market.
  1. Macroeconomic environments (business cycles)
  2. Technological progress across industries (PCs, industrial robots, AIs)

- Skills may be over-supplied or under-supplied in labour market

★ Conceptually well known, but literature on measurement of mismatch is thin

This paper

- Quantify college-major-occupation mismatch
- What we do:
  1. We propose an accounting exercise of skill mismatch
     - Estimate college-major returns by occupation
     - Correct biases from selection into an occupation (and major [WIP])
     - Accounting based on the wage penalty for not working in the *best-match*
  2. General equilibrium [WIP]
  3. Use ATO's ALife merged with Dept. of Education's HECS-HELP
     - Labour market histories
     - Field of study (college majors)
     - Australia: university finance scheme providing exogenous variation for study selection

Literature

- Mesurement of Mismatch: Pellizzari and Fichen (2013), McGowan and Andrews (2015), Guvenen, Kuruscu, Tanaka and Wiczer (2020), Lise and Postel-Vinay (2020), Fredriksson et al. (2018), Groes et al. (2015), Baley et al. (2022)
- Task approach: Autor et al. (2003), Ingram and Neumann (2006), Poletaev and Robinson (2008), Gathmann and Schonberg (2010), Bacolod and Blum (2010), Yamaguchi (2012), Rendall (2010)
- Occupational selection: Roy (1951), Heckman and Sedlacek (1985), Lee (1983), Dahl (2002), Eckardt (2019)
- Returns to college majors: Altonji et al. (2012), Altonji et al. (2016b), Altonji et al. (2014), Altonji et al. (2016a), Kirkeboen et al. (2016), Liu et al. (2016)

Outline

Baseline Model

Consider a unit mass of individual workers $i \in [0, 1]$

- Individual i studies major j at college: $h_i^j = 1$, otherwise $h_i^j = 0$.
- After graduation, every year t, individual chooses and works in some occupation k.
- Individual i with major j and characteristics $X_{i,t}$ supplies major-specific skill to the markets,

$$\tilde{h}_i^j = h_i^j e^{\gamma X_{i,t} + \lambda_{r,t} + \nu_c}$$

where $\lambda_{r,t}$ is a region-time-specific effect and ν_c is a cohort-specific effect

Baseline Model (cont'd)

A worker's wage is determined by $\tilde{h}_i^j$ and prices of skills $\{p_j^k\}_{j,k}$ at different occupations

- Prices potentially differ across occupations $(p_j^k \neq p_j^{k'})$. Due to, e.g., labour market segmentation.
- Wage is given by

$$w_{i,j,k,c,r,t} = p_j^k \tilde{h}_i^j = p_j^k e^{\gamma X_{i,t} + \lambda_{r,t} + \nu_c}$$

- Base regression equation is,

$$\ln w_{i,j,k,c,r,t} = \underbrace{\ln\left(p_j^k\right)}_{\beta_{j,k} D_{i,(k|j),c,r,t}} + \gamma X_{i,t} + \lambda_{r,t} + \nu_c + \epsilon_{i,j,k,c,r,t}$$

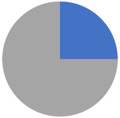
What is Mismatch?

Field of Study
Options

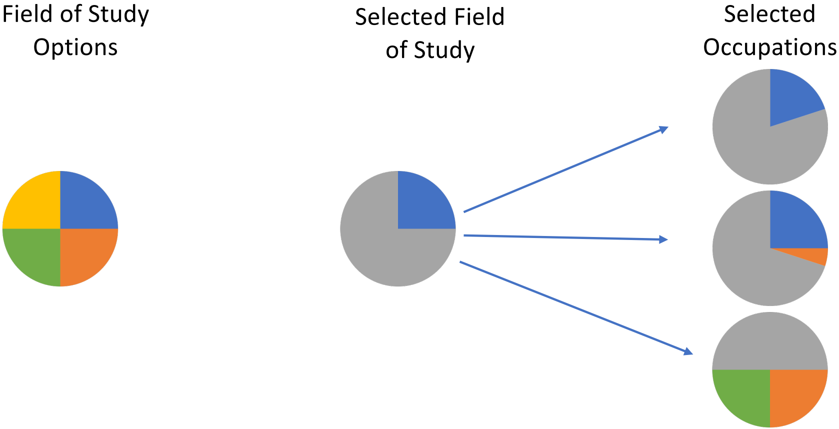
What is Mismatch?

Field of Study Options

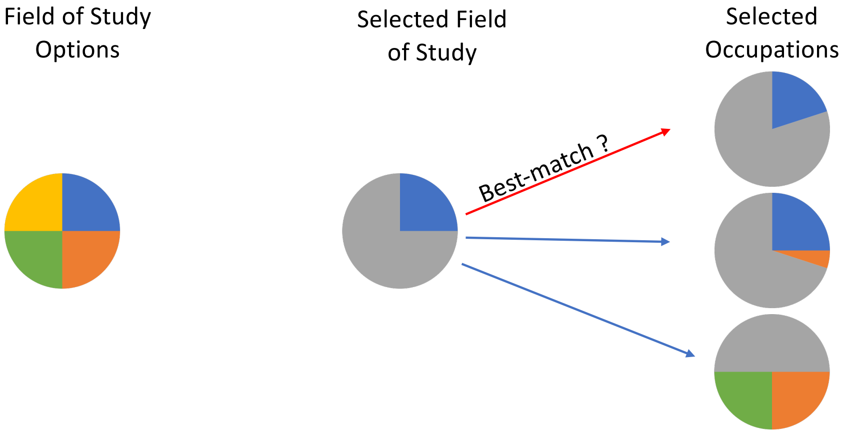
Selected Field of Study



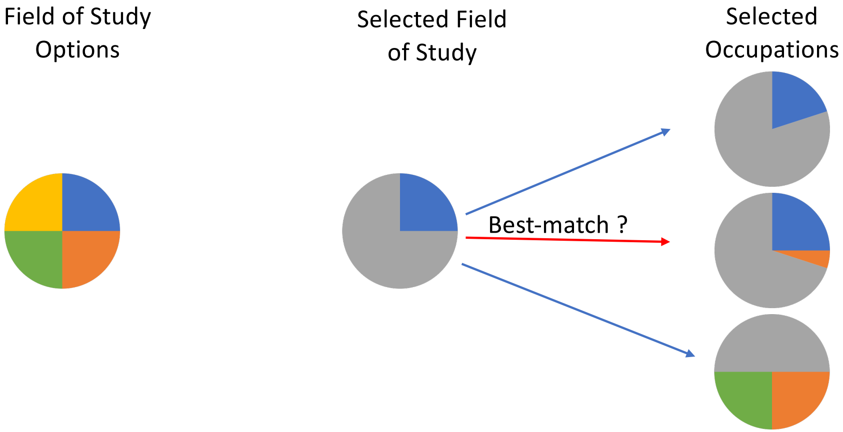
What is Mismatch?



What is Mismatch?



What is Mismatch?



Baseline Mismatch Measure

- Wage penalty for not working at the best-match k^* given major j is,

$$\zeta_{(k|j)} = \beta_{j,k^*} - \beta_{j,k}.$$

- Average per-capita, mismatch penalty for each major j,

$$m_j = \sum_{k \in \mathcal{K}} \zeta_{(k|j)} \frac{n_{j,k}}{\sum_{k \in \mathcal{K}} n_{j,k}}$$

- Aggregate mismatch,

$$\mathcal{M} = \sum_{j \in \mathcal{J}} m_j \frac{\sum_{k \in \mathcal{K}} n_{j,k}}{\sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{J}} n_{j,k}}$$

★ Economy's gain by allocating workers into best-match in partial equilibrium - fixed $\beta_{j,k}$

General Equilibrium

- Assume a production function for each occupation

$$F^k(H_1^k, ..., H_j^k, ..., H_J^k) = A^k \left(\alpha_1^k \left(H_1^k\right)^{\sigma_k} + ...\alpha_j^k \left(H_j^k\right)^{\sigma_k} + ...\alpha_J^k \left(H_J^k\right)^{\sigma^k} \right)^{\frac{1}{\sigma_k}}$$

where $H_j^k = \sum_{i \in I^k} \tilde{h}_i^j$

- From FOC, we have

$$\frac{p_{j'}^k}{p_j^k} = \frac{\partial F^k / \partial H_{j'}^k}{\partial F^k / \partial H_j^k} = \frac{\alpha_{j'}^k}{\alpha_j^k} \left(\frac{H_{j'}^k}{H_j^k} \right)^{\sigma^k - 1}$$

- By taking logarithm on both sides,

$$\underbrace{\ln\left(p_{j'}^k / p_j^k\right)}_{\beta_{j'}^k - \beta_j^k} = \ln\left(\alpha_{j'}^k / \alpha_j^k\right) + \left(\sigma^k - 1\right) \ln\left(H_{j'}^k / H_j^k\right)$$

- We estimate this equation with time-region observations

General Equilibrium (cont'd)

- We can recover $\left\{A^k\right\}_k$ from

$$\frac{p_j^{k'}}{p_j^k}=\frac{A^{k'}\left(\alpha_1^{k'}\left(H_1^{k'}\right)^{\sigma_{k'}}+\ldots+\alpha_j^{k'}\left(H_j^{k'}\right)^{\sigma^{k'}}\right)^{\frac{1}{\sigma^{k'}}-1}\alpha_j^{k'}\sigma^{k'}\left(H_{j'}^{k}\right)^{\sigma^{k'}-1}}{A^k\left(\alpha_1^k\left(H_1^k\right)^{\sigma_k}+\ldots+\alpha_j^k\left(H_j^k\right)^{\sigma^k}\right)^{\frac{1}{\sigma^k}-1}\alpha_j^k\sigma^k\left(H_j^k\right)^{\sigma^k-1}}$$

where $\frac{p_j^{k'}}{p_j^k}=\frac{\partial F^{k'}/\partial H_j^{k'}}{\partial F^k/\partial H_j^k}$

- Define $\left(\hat{H}_1^k,\ldots,\hat{H}_j^k,\ldots\hat{H}_J^k\right)$ as the optimal allocation ($K\times J$ unknowns). They are solutions of the following equations ($(K-1)\times J+J=K\times J$).

$$\frac{\partial F^{k'}/\partial\hat{H}_j^{k'}}{\partial F^k/\partial\hat{H}_j^k}=1.$$

$$\hat{H}_j^1+\ldots+\hat{H}_j^k+\ldots+\hat{H}_j^K=\bar{H}_j,\quad\forall j\in J$$

General Equilibrium Mismatch

- Mismatch for major j,

$$m_j = \frac{1}{n} \sum_k \left(F^k \left(\hat{H}_1^k, ..., \hat{H}_j^k, ...\hat{H}_J^k \right) - F^k \left(\hat{H}_1^k, ..., H_j^k, ...\hat{H}_J^k \right) \right)$$

- Aggregate mismatch,

$$\mathcal{M} = \frac{1}{n} \sum_k \left(F^k \left(\hat{H}_1^k, ..., \hat{H}_j^k, ...\hat{H}_J^k \right) - F^k \left(H_1^k, ..., H_j^k, ..., H_J^k \right) \right)$$

Outline

Selection Bias

We potentially have two selection issues

1. Selection into an occupation
   - Error term might be correlated with occupational choices
2. Selection into a major
   - Error term might be correlated with college major choices

Today's result, only controlling for 1.

Control Function Approach

Dahl (2002) method to correct for sample selection bias in (j,k) cells

- Earnings model becomes,

$$\ln w_{i,j,k,r,t} = \sum_{j'}^{J}\sum_{k'}^{K}\beta_{j',k'}D_{i,k'|j',c,r,t} + X_{i,t}\gamma + \lambda_{r,t} + \nu_c + \rho_{j,k}\cdot g(\hat{p}_{i,(k|j),c,r,t}) + \xi_{i,j,k,c,r,t}$$

$$\text{where } E[\xi_{i,j,k,r,t}|M_{i,(k|j),c,r} = 1] = 0$$

  - Every period, individuals make occupational choices based on observed/unobserved earnings ability and non-monetary utility in an occupation.

1. Estimate $p_{i,(k|j),c,r,t}$ with occupation shift-share (Bartik) instrument
2. Construct parametric control function, $g(\hat{p}_{i,(k|j),c,r,t})$ (Lee, 1983).

Outline

ALife Data by ATO

- ALife: personal income tax records for a 10% representative sample
  - Use yearly wages for each individual
- Merge dataset with Department of Education study records
  - Use field of study (major) at college, and the year of graduation

Sample section

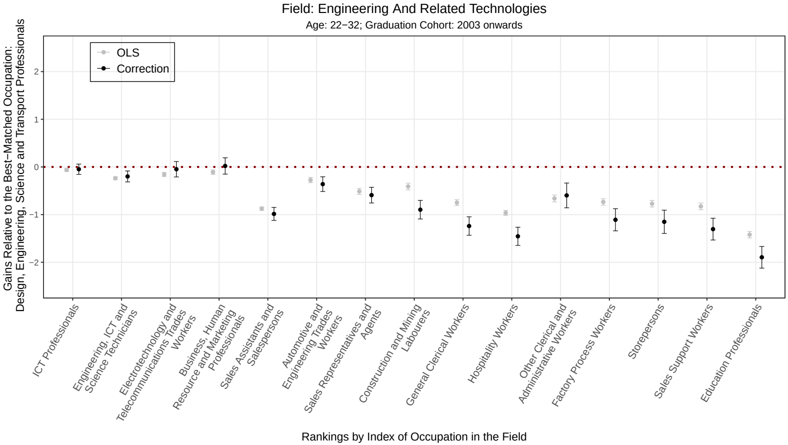
- Restrict our sample to individuals aged 22 to 32
- Starts from the cohort who graduated in 2003 (b/c of Bartik)
- Remove observations in managerial occupations (Kambourov and Manovskii, 2009)

Summary Statistics: Occupations by Field of Study (Major)

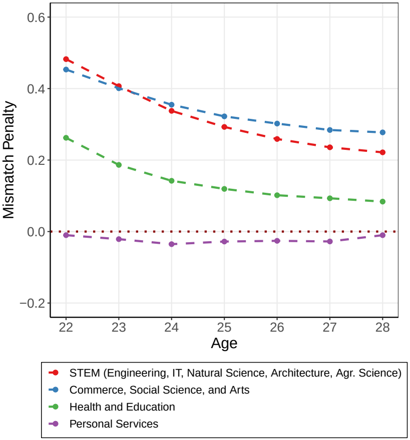
	Occ. Shares			No.Occ.
	Mean	Min	Max	
Natural And Physical Sciences	0.043	0.011	0.270	21
Information Technology	0.045	0.007	0.422	21
Engineering And Related Technologies	0.038	0.004	0.504	26
Architecture And Building	0.059	0.013	0.338	15
Agriculture, Environmental And Related Studies	0.072	0.023	0.266	10
Health	0.032	0.002	0.584	31
Education	0.035	0.002	0.634	28
Management And Commerce	0.028	0.003	0.351	36
Society And Culture	0.028	0.003	0.182	36
Creative Arts	0.029	0.004	0.139	34
Food, Hospitality And Personal Services	0.086	0.035	0.188	9

Outline

Estimated Major-Occupation Mismatch: Engineering

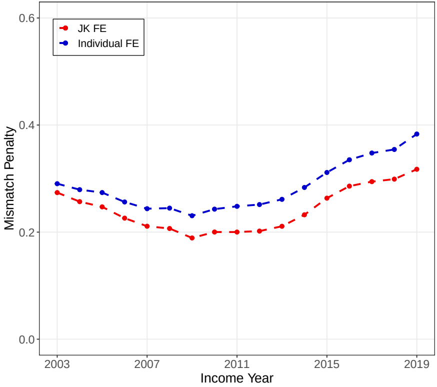


Average Occupation Mismatch over the Life-Cycle



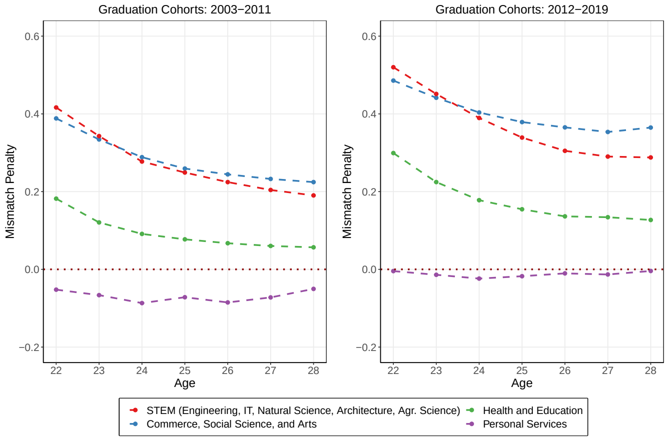
- STEM, Commerce and Social Science, Arts majors: high, but "steep" decline of mismatch over the life-cycle.
- Health and Education/Other majors: low and flatter mismatch.

Evolution of Aggregated Occupational Mismatch



- Increasing from about 2009-2011.

Average Mismatch over the Life-Cycle by Cohorts



- 2012 onward: larger and more persistent mismatch.
- Removal of caps on Commonwealth Supported Places (CSPs) $\Rightarrow$ Large supply side shock with higher mismatch

Decomposition of Skill Mismatch by Field of Study

	Group of Fields of Study			
O*NET Skills	STEM	Commerce/Social/Arts	Health/Education	Personal Services
Engineering and Technology	0.262	0.143	-0.027	-0.038
Mathematics and Science	0.100	0.038	-0.008	0.000
Health Services	-0.031	0.004	0.158	-0.011
Manufacturing and Production	-0.029	-0.043	0.004	0.059
Law and Public Safety	-0.004	0.190	-0.008	-0.003
Business and Management	-0.010	0.028	-0.005	-0.025
Art and Humanities	-0.009	-0.017	0.101	-0.017
Communications	-0.033	0.041	-0.046	-0.105
Education	-0.001	-0.001	0.018	0.000
Transportation	-0.043	-0.089	-0.079	0.241
Overall Skill Mismatch	**0.203**	**0.293**	**0.108**	**0.101**

Notes: Decomposition of occupational skill mismatch penalty for each major. Column (1) lists 10 clusters of skills classified by O*NET. Columns (2)-(4) decomposition of overall skill mismatch. Last row overall occupational skill mismatch for each major.

Summary and Future Work

This paper:

- Accounting exercise of college-major-occupation mismatch
- Apply it to ALife merged with HECS-HELP information
- We find:
  - Large mismatch for STEM, Commerce and Social Science majors
  - Negative impact of education subsidies (preliminary)

Future work:

- Complete accounting exercise of general equilibrium mismatch and optimal allocation
- Exploring the causes of mismatch in skills (language, math, etc.)
- Dynamic model
- Policy analyses