- Heckman (1979) outlines the canonical sample selection problem

- Heckman (1979) outlines the canonical sample selection problem
- e.g. we only observe the earnings of individuals who are employed

- Heckman (1979) outlines the canonical sample selection problem
- e.g. we only observe the earnings of individuals who are employed
- This might distort our estimates of wage returns to skill

- Heckman (1979) outlines the canonical sample selection problem
- e.g. we only observe the earnings of individuals who are employed
- This might distort our estimates of wage returns to skill
- Can we improve on this by using machine learning?

- Heckman (1979) outlines the canonical sample selection problem
- e.g. we only observe the earnings of individuals who are employed
- This might distort our estimates of wage returns to skill
- Can we improve on this by using machine learning?
  - Especially if the choice dimension is much larger than work/not work?

Heckman selection correction two-step procedure:

Heckman selection correction two-step procedure:

- Step 1: Model the selection decision (e.g., work vs. not work)

Heckman selection correction two-step procedure:

- Step 1: Model the selection decision (e.g., work vs. not work)
    - Estimate a probit model to get selection probabilities

Heckman selection correction two-step procedure:

- Step 1: Model the selection decision (e.g., work vs. not work)
  - Estimate a probit model to get selection probabilities
  - Compute the inverse Mills ratio $\lambda(\cdot)$

Heckman selection correction two-step procedure:

- Step 1: Model the selection decision (e.g., work vs. not work)
    - Estimate a probit model to get selection probabilities
    - Compute the inverse Mills ratio $\lambda(\cdot)$
- Step 2: Include $\lambda(\cdot)$ as a regressor in the wage equation

Heckman selection correction two-step procedure:

- Step 1: Model the selection decision (e.g., work vs. not work)
  - Estimate a probit model to get selection probabilities
  - Compute the inverse Mills ratio $\lambda(\cdot)$
- Step 2: Include $\lambda(\cdot)$ as a regressor in the wage equation

$$\log \mathit{wage} = X\beta + \lambda(Z\gamma) + u$$

Heckman selection correction two-step procedure:

- Step 1: Model the selection decision (e.g., work vs. not work)
  - Estimate a probit model to get selection probabilities
  - Compute the inverse Mills ratio $\lambda(\cdot)$
- Step 2: Include $\lambda(\cdot)$ as a regressor in the wage equation

$$\log wage = X\beta + \boxed{\lambda(Z\gamma)} + u$$

- The $\lambda(\cdot)$ term corrects for selection bias

Limitations of the Heckman approach

Limitations of the Heckman approach

- Requires normality assumption for the error terms

Limitations of the Heckman approach

- Requires normality assumption for the error terms
- Need Z's that affect selection but not the outcome (exclusion restriction)

Limitations of the Heckman approach

- Requires normality assumption for the error terms
- Need Z's that affect selection but not the outcome (exclusion restriction)
- Difficult to extend to multinomial choice settings

Limitations of the Heckman approach

- Requires normality assumption for the error terms
- Need Z's that affect selection but not the outcome (exclusion restriction)
- Difficult to extend to multinomial choice settings
  - With many alternatives, the selection correction becomes intractable

Annals of Economics and Statistics, Number 142, June 2021

SELECTIVE MIGRATION, OCCUPATIONAL CHOICE, AND THE WAGE RETURNS TO COLLEGE MAJORS

TYLER RANSOM[a]

I examine the extent to which the monetary returns to college majors are influenced by selective migration and occupational choice across locations in the US. To quantify the role of selection, I develop and estimate an extended Roy model of migration, occupational choice, and earnings where, upon completing their education, individuals choose a location in which to live and an occupation in which to work. In order to estimate this high-dimensional choice model, I make use of machine learning methods that allow for model selection and estimation simultaneously in a non-parametric setting. I find that OLS estimates of the return to business and STEM majors relative to education majors are biased upward by 15% at the median. Selection is strongest in locations in the Northeastern US.

JEL Codes: I2, J3, R1.

Keywords: College Major, Migration, Occupation, Roy Model.

1. INTRODUCTION

One lesser-known characteristic of the US labor market is that the wage returns to different college majors are highly heterogeneous across space. For example, among men in the 2010–2019 American Community Survey, the return to STEM and business majors each range from about 23% to over 47%, relative to education majors.[1] While much work has examined sorting of majors into occupations, occupational sorting does little to narrow this gap: the return to a STEM major in a STEM occupation relative to a STEM major in a non-STEM occupation ranges from 11% in Pennsylvania to 43% in Washington, DC, with a similar range for other majors. This broad range in returns to majors and occupations suggests that post-college migration, and in particular its interaction with post-college occupational choice, might be a significant driver of the observed spatial variation in earnings.

The objective of this paper is to uncover the extent to which selection into residence location and occupation biases the observed monetary returns to college majors (relative to education majors). Aside from Winters (2017), this is the first paper to examine the spatial dimension of college major and occupation decisions.[2]

Understanding the true returns to human capital investments is important because students base these investment decisions in part on expected earnings (Beffy, Fougère, and Maurel, 2012; Wiswall and Zafar, 2015). A student might choose differently if an ob-

I would like to thank Peter Arcidiacono, Esteban Aucejo, Pat Bayer, Rob Garlick, Arnaud Maurel, Ronni Pavan, Kevin Stange and various conference and seminar participants for their helpful discussions and comments. Special thanks to Jamin Speer for generously providing helpful code for classifying college majors in the ACS. All errors are my own. Full replication code is available at DOI 10.5281/zenodo.4907446.

[a]Department of Economics, University of Oklahoma, 308 Cate Center Dr. 158 CCD1, Norman, Oklahoma, USA 73072. `ransom@ou.edu`

[1]Returns calculated using a Mincerian regression of log earnings on a cubic in potential experience, demographic indicators, and MSA fixed effects.

[2]Winters (2017) examines the migration response of different college majors to birth-state earnings shocks to workers in the same major.

Geographic heterogeneity in wage returns to college major

Geographic heterogeneity in wage returns to college major

- Individuals choose where they live based on wages and non-wage factors

Geographic heterogeneity in wage returns to college major

- Individuals choose where they live based on wages and non-wage factors
- Problem: researcher only sees wages in chosen residence location

Geographic heterogeneity in wage returns to college major

- Individuals choose where they live based on wages and non-wage factors
- Problem: researcher only sees wages in chosen residence location
- Thus, wage returns are potentially contaminated by selection bias

Heckman model: the inverse Mill's ratio $\lambda(\cdot)$ corrects for selection

Heckman model: the inverse Mill's ratio $\lambda(\cdot)$ corrects for selection

$$\log \mathit{wage} = X\beta + \lambda\left(Z\gamma\right) + u$$

Heckman model: the inverse Mill's ratio $\lambda(\cdot)$ corrects for selection

$$\log \mathit{wage} = X\beta + \lambda\left(Z\gamma\right) + u$$

One can generalize this approach to multinomial choice and non-normality

Heckman model: the inverse Mill's ratio $\lambda(\cdot)$ corrects for selection

$$\log wage = X\beta + \lambda\left(Z\gamma\right) + u$$

One can generalize this approach to multinomial choice and non-normality

$$\log wage = X\beta + \sum_j d_j \widetilde{\lambda}\left(p_j(Z), p_k(Z)\right) + u$$

Heckman model: the inverse Mill's ratio $\lambda(\cdot)$ corrects for selection

$$\log wage = X\beta + \lambda\left(Z\gamma\right) + u$$

One can generalize this approach to multinomial choice and non-normality

$$\log wage = X\beta + \sum_j d_j \widetilde{\lambda}\left(p_j(Z), p_k(Z)\right) + u$$

where

- d_j is a dummy for living in location j

Heckman model: the inverse Mill's ratio $\lambda(\cdot)$ corrects for selection

$$\log wage = X\beta + \lambda\left(Z\gamma\right) + u$$

One can generalize this approach to multinomial choice and non-normality

$$\log wage = X\beta + \sum_{j} d_j \widetilde{\lambda}\left(p_j(Z), p_k(Z)\right) + u$$

where

- d_j is a dummy for living in location j
- $\widetilde{\lambda}$ is a flexible function

Heckman model: the inverse Mill's ratio $\lambda(\cdot)$ corrects for selection

$$\log wage = X\beta + \lambda\left(Z\gamma\right) + u$$

One can generalize this approach to multinomial choice and non-normality

$$\log wage = X\beta + \sum_j d_j \widetilde{\lambda}\left(p_j(Z), p_k(Z)\right) + u$$

where

- d_j is a dummy for living in location j
- $\widetilde{\lambda}$ is a flexible function
- p_j and p_k are probabilities of choosing j or k (as a function of Z)

The p's on the previous slide are selection probabilities

The p's on the previous slide are selection probabilities

- p_j is the probability of choosing the chosen alternative

The p's on the previous slide are selection probabilities

- p_j is the probability of choosing the chosen alternative
- p_k is the probability of choosing the next-preferred alternative

The p's on the previous slide are selection probabilities

- p_j is the probability of choosing the chosen alternative
- p_k is the probability of choosing the next-preferred alternative
- Use a classification tree model to obtain the p's

The p's on the previous slide are selection probabilities

- p_j is the probability of choosing the chosen alternative
- p_k is the probability of choosing the next-preferred alternative
- Use a classification tree model to obtain the p's
- Assume that individuals with same values of Z and similar p's have identical tastes

This approach improves on a bin-estimation approach (see Dahl, 2002, *Econometrica*)

This approach improves on a bin-estimation approach (see Dahl, 2002, *Econometrica*)

- Can include a higher dimension of Z while limiting the curse of dimensionality

This approach improves on a bin-estimation approach (see Dahl, 2002, *Econometrica*)

- Can include a higher dimension of Z while limiting the curse of dimensionality
- Tree models adaptively partition the covariate space

This approach improves on a bin-estimation approach (see Dahl, 2002, *Econometrica*)

- Can include a higher dimension of Z while limiting the curse of dimensionality
- Tree models adaptively partition the covariate space
- Avoids the need to manually specify bins or interactions