

Literature

Adrien Foutelet

February 10, 2025

1 Model

- What is the research question?
 - How does the research question relate to existing theoretical and empirical literature?
 - Why is it worth asking?
- What are the data being used here?
 - How were they collected?
 - What are the important variables?
 - How are they defined?
 - What is the unit of observation?
- What is the empirical strategy for answering this research question?
 - What would the ideal data set look like? What empirical strategy would you use on it?
 - How is the data set in this paper different from that ideal data set?
 - How does identification work in this paper?
 - What are the sources of exogenous variation?
- What econometric techniques are being used in this paper?
 - Are they appropriate?
 - What is the central estimating equation (or equations)?
 - What is assumed to be exogenous and endogenous? What is in the unobservable term?
 - Are you worried about any of the assumptions that are needed for identification?
- What are the main conclusions of the author?
- What alternative interpretations of the results are plausible?
 - Does the author test his/her conclusions against alternative interpretations?
 - Does the author provide any practical reason why these other interpretations are less likely?

2 ECON2340 Labor Economics II

2.1 The canonical model of skill differentials

2.2 Skills, tasks, and technologies

2.3 Superstars and Mediocrities

2.4 Comparative advantage, self-selection, and the Roy model

2.4.1 Borjas, George J. “Self-Selection and the Earnings of Immigrants.” *The American Economic Review*, vol. 77, no. 4, 1987, pp. 531–53.

Research question: How are the earnings of the immigrant population compared to those of the native population affected by the endogeneity of the decision to migrate?

2.4.2 Chandra, Amitabh, and Douglas O. Staiger. “Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks.” *Journal of Political Economy* 115.1 (2007): 103-140.

2.4.3 Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad. “Field of Study, Earnings, and Self-Selection.” *The Quarterly Journal of Economics* 131.3 (2016): 1057-1111.

2.4.4 Hanson, Gordon, and Chen Liu. “Immigration and occupational comparative advantage.” *Journal of International Economics* 145 (2023): 103809.

- What is the research question?

How do high-skilled immigrants sort into occupations in the U.S., and to what extent is this sorting driven by their origin-country’s education quality and linguistic similarity to English?

- How does the research question relate to existing theoretical and empirical literature?

Comparative advantage and labor sorting: it follows the Roy model of occupational choice, which predicts that workers select jobs where they have the highest productivity (Roy, 1951); it aligns with studies on task-based labor economics (Autor et al., 2003) and skill-biased sorting (Costinot and Vogel, 2010).

Immigration and occupational specialization: prior research documents occupational clustering of immigrants by country of origin (Hanson and Liu, 2017; Patel and Vella, 2013); the paper adds to this by quantifying how education quality (pisa scores) and linguistic similarity (ASJP metric) influence occupational choice.

Migration networks and selection: it extends work on migration costs and network effects (Munshi, 2003; Beine et al. 2008) by showing that comparative advantage persists even after controlling for network spillovers;

it contrasts with studies on occupational downgrading (Dustmann et al., 2013), showing that high-skilled immigrants often sort efficiently rather than experiencing systematic skill mismatches.

Immigration policy and skill selection: connects with work on immigration policy effects (Borjas, 1993; Kaushal and Lu, 2015) by comparing the US and Canada; supports the argument that origin-country education matters more than visa allocation mechanisms in explaining occupational sorting.

- Why is it worth asking?

Immigration is a very controversial issues in developed countries these days.

- What are the data being used here?

Employment and demographics: US American community survey, CA National household survey.

Skill and education quality: PISA; ASJP-based phonetic similarity measure.

Occupational characteristics: O*NET and Dictionary of occupational titles for task intensities.

Migration costs and controls: CEPII bilateral distance data; country-of-origin and occupation fixed effects to account for unobserved heterogeneity.

Sample restriction: college-educated, prime age, men.

- How were they collected?
- What are the important variables?

Dependent variable: occupational employment share + log share of hours worked by immigrants from country s in occupation o in destination d .

Independent variables: cognitive skill of s ; linguistic proximity of s ; occupational task intensities of o (cognitive reasoning intensity, communication intensity, routine task intensity, manual task intensity); interaction terms (PISA times cognitive reasoning intensity to test if workers from high-PISA countries concentrate in cognitive-intensive occupations, linguistic proximity times communication intensity to test if workers from linguistically similar countries concentrate in communication-intensive occupations).

Control variables: geographic distance from the origin country to the US; country fixed effects (for migration costs and alternative employment); occupational fixed effects (average wages and job requirements in each occupation).

- How are they defined?
- What is the unit of observation?

The employment share of college-educated, prime-age male immigrants from a given origin country s in a specific occupation o within a destination country d .

- What is the empirical strategy for answering this research question?

Fréchet-Roy occupational choice model and double-differencing regression, to estimate how high-skilled immigrants sort into occupations based on comparative advantage.

- What would the ideal data set look like? What empirical strategy would you use on it?

We wish we also had pre-migration occupation and work experience, language proficiency test scores, visa type and entry mechanism, precise age at migration, more detailed occupation classification, job posting and skill requirements, earnings and promotion trajectory, panel data tracking immigrants over time (occupational changes, skill downgrading effects).

Exogenous policy changes (visa lotteries) to test for causality in sorting patterns.

Randomized work permit allocations to observe how employment choices differ with and without work restrictions.

- How is the data set in this paper different from that ideal data set?

Causality? IV? Use historical migration patterns or past colonial ties as instruments for migration costs, disentangling comparative advantage from migration network effects.

- How does identification work in this paper?

Comparative Advantage Hypothesis: Immigrants sort into occupations where they have relative productivity advantages.

Double Differencing: Neutralizes the effects of occupation-specific wages and country-specific migration factors.

Age-at-Arrival Analysis: Immigrants who arrived after age 18 (educated abroad) should exhibit stronger sorting patterns than those who arrived younger (educated in the U.S.).

Comparison with Canada: If similar sorting patterns exist despite different immigration policies, results reflect education quality and linguistic proximity rather than U.S. visa selection biases.

Some countries might push migrants to go for specific professions.

- What are the sources of exogenous variation?

Origin-country education quality, linguistic proximity, age-at-arrival analysis (exposure to US education versus origin-country education).

Comparison with Canada: Canada's skill-based immigration policy is an exogenous benchmark to test if sorting patterns arise from education quality rather than US visa policies.

- What econometric techniques are being used in this paper?

- Are they appropriate?
- What is the central estimating equation (or equations)?
- What is assumed to be exogenous and endogenous? What is in the unobservable term?
- Are you worried about any of the assumptions that are needed for identification?

- What are the main conclusions of the author?

Immigrant occupational sorting is driven by comparative advantage: High-skilled immigrants specialize in jobs that align with their origin-country education quality (PISA scores) and linguistic proximity to English.

Cognitive-intensive jobs attract workers from high-PISA countries, while communication-intensive jobs attract workers from linguistically similar countries.

Education quality, not U.S. immigration policy, explains sorting patterns, as similar trends are observed in Canada.

Age-at-arrival matters: Immigrants who arrived after age 18 (educated abroad) show strong sorting, while those who arrived younger do not, confirming that origin-country education, not cultural preferences, drives specialization.

Policy implication: A shift toward skill-based immigration (like Canada's system) would likely increase STEM immigration from high-PISA countries (East Asia, Europe) and reduce inflows from lower-scoring regions (Latin America, Africa).

- What alternative interpretations of the results are plausible?

Origin country immigration policy, cultural and social norms, network effects (they put an AR-A process).

Lorenzo: They don't have measures of routine in tasks in the origin countries; they have type of college degrees and don't seem to utilize them properly (they only do it with people having computer degrees).

- Does the author test his/her conclusions against alternative interpretations?
- Does the author provide any practical reason why these other interpretations are less likely?

References

- [1] G. Chodorow-Reich, M. Smith, O. Zidar, E. Zwick. Tax Policy and Investment in a Global Economy. Working Paper, 2024.