Bias-Corrected Changes in the Wage Equation of In-Migrants Following COVID-19

Labor Economics Week 4 Assignment

Steven VanOmmeren
Boston College

April 6, 2025

Overview

Introduction

- A non-trivial percent of Americans move across states each year (in-migration). Why?
- **Roy model**: people migrate if they expect to make more money at the new location (after accounting for moving costs)
- Problem: because migration is non-random, this imposes a selection bias when analyzing immigrant wages
- To control for in-migration bias, we need to model the probability of moving, then incorporate a correction term into our model of wages. This is called a **Heckit** procedure (due to Heckman).
- Finally, because COVID-19 may have permanently affected the wage factors of in-migrants (e.g. increased work from home), I compare pre-COVID and post-COVID data results as well.

Methods: Data

- To study the wages of U.S. in-migrants, I use data from the American Community Survey (ACS).
- Two outcome variables of interest: whether the person moved in the past year, and log of real wages:
  - flag_inmigrate: indicator for MIGRATE1 == 3 ("Moved between states")
  - log(wage): log(INCWAGE_CPIU_2010 / (UHRSWORK * WKSWORK2))
  - wages are deflated using the CPI-U to 2010 values
- Data were collected from 2017 to 2023 for all states, yielding a large dataset with many factors to analyze.

* WKSWORK2 was transformed from an intervalled variable to a numeric count of weeks worked in the last year.

Methods: Data Filters

- The following filters were used to isolate the wage effects of in-migration, while excluding international migration decisions:
    - Limit to head of household
    - Age: 18-65
    - Limit to employed wage earners
    - Limit to natural-born U.S. citizens
    - Must have lived in a U.S. state last year
    - Real income $\geq \$200,000$ is considered an outlier
    - Real wage below \$5/hr or above \$100/hr is considered an outlier
    - Remove self-employed individuals (following Borjas (1987))
- This yields 3,165,983 observations, with the sample representing **339,242,941 person-years** using ACS person weights.*

* I aggregated over my control variables of interest to reduce computation time, summing up the ACS person weights. My regression dataset has 507,255 observations.

Methods: Analysis

- I test 4 models of wages on in-migrants in the data:
  1. **Uncorrected**: OLS regression of log(wage) on control variables using all years.
  2. **Corrected**: Heckit-corrected regression of log(wage) on control variables using all years.
  3. **Pre-COVID**: Heckit-corrected regression of log(wage) on control variables using 2017-2019 data.
  4. **Post-COVID**: Heckit-corrected regression of log(wage) on control variables using 2020-2023 data.
- Models 2-4 include the Inverse Mills Ratio from a corresponding probit regression of in-migrant status on control variables (with corresponding year filters).
- Control variables for log(wage) models: sex, age, number of children, marriage status, veteran status, race, education, spouse characteristics (employed, college-educated, veteran), female interaction terms, and year fixed effects.
- Control variables for in-migrant status models: sex, age, number of children, marriage status, veteran status, single mother status, race, education, work from home status, health insurance via employer status, spouse characteristics (employed, college-educated, veteran), lives in metropolitan area, year fixed effects, and state fixed effects.

Results: Summary Statistics by In-migrant Status

Probability of moving across states in the past year: 2.7%.

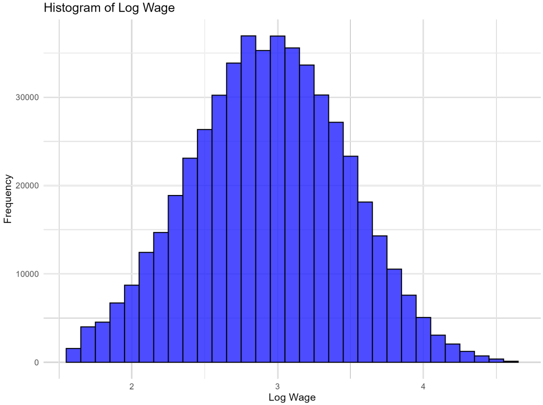
Compared to stayers, in-migrants:

- are substantially younger
- have fewer children
- have fewer females
- have less household income
- are more likely to be veterans
- are more educated

	In-migrant Status	
Statistic	Stayers	Movers
Observations	3,090,402	75,581
Represented Person-Years	330,170,014	9,072,927
Average real wage	22.58	22.24
Average real household income	75,015	66,980
Average real wage income	47,814	46,263
Average # of children	0.82	0.47
Average hours worked per week	41.40	42.17
Average age	43.01	34.72
Percentage Asian	2.23	4.01
Percentage White	80.80	82.28
Percentage Black	14.71	13.40
Percentage female	50.55	45.88
Percentage veteran	5.94	7.49
Percentage with less than high school education	3.58	1.92
Percentage with post-college education	14.59	21.09
Percentage with employed spouse	44.13	33.65
Percentage of power couples	28.12	29.37

Results: Distribution of Log Wages

After filtering the data, log wages are approximately Normally distributed.



Results: Selection Probit Models

Probit models confirm that:

- younger people are far more likely to move
- compared to Asian and White people, Black people are less likely to relocate (though slightly more likely to move than omitted category "Other")
- spouse characteristics matter:
  - More likely to move when spouse is more educated
  - More likely to move when spouse is a veteran
  - Less likely to move when spouse is employed
- people with health insurance through their employer are less likely to move
- remote workers are more likely to relocate
- Pre- and Post-COVID effects are similar. Post-COVID, remote workers are less likely to relocate than Pre-COVID. Very young people are slightly more likely to move following COVID.

	Dependent Variable: In-migrant Status		
	Full Sample	Pre-COVID	Post-COVID
Age 18-24	0.92***	0.90***	0.93***
Age 25-34	0.61***	0.62***	0.60***
Age 35-44	0.38***	0.40***	0.36***
# Children	−0.27***	−0.28***	−0.26***
# Children Under 5 Yrs	0.06***	0.07***	0.05***
Female	−0.05***	−0.04***	−0.05***
Ever Married	0.13***	0.15***	0.13***
Veteran	0.19***	0.21***	0.17***
Single Mother	0.03***	0.03***	0.04***
Asian	0.12***	0.10***	0.13***
White	0.09***	0.09***	0.09***
Black	0.02***	0.04***	0.01***
Less than High School	−0.14***	−0.16***	−0.12***
Post-College Education	0.23***	0.24***	0.22***
Works from Home	0.21***	0.29***	0.19***
Health Insurance through Work	−0.15***	−0.15***	−0.14***
Employed Spouse	−0.24***	−0.25***	−0.23***
College-Educated Spouse	0.15***	0.16***	0.14***
Veteran Spouse	0.10***	0.09***	0.11***
Lives in Metro. Area	0.11***	0.11***	0.11***
Constant	−2.33***	−2.38***	−2.35***
Year Fixed effects	*Yes*	*Yes*	*Yes*
State Fixed effects	*Yes*	*Yes*	*Yes*
Observations	507,255	213,107	294,148
ρ	0.07	0.38	-0.15
Inverse Mills Ratio	0.03*(0.01)	0.21***(0.02)	-0.08***(0.02)
Note:	*p<0.05; **p<0.01; ***p<0.005		

Results: Wage Regressions (Uncorrected and Bias-Corrected)

In-migrant wages:

- While the IMR coefficient is significant ($p < 0.05$), selection bias makes little to no practical difference to the coefficients (Uncorrected vs Corrected).
- The sign of the IMR coefficient flips between Pre-COVID and Post-COVID models. Both are small but highly significant.
- There are significant gender pay gaps in all models
- Post-COVID wages:
  - Have smaller gender gap than Pre-COVID
  - Young workers make comparatively less across the board (vs 45-65 year old workers)
  - Women with older children earn comparatively more than Pre-COVID
  - Increased education has a smaller wage effect than Pre-COVID
  - Earn comparatively more than Pre-COVID when spouse is employed

	Dependent Variable: Log(Wage)			
	Uncorrected	Corrected	Pre-COVID	Post-COVID
Female	-0.12^{***}	-0.12^{***}	-0.16^{***}	-0.09^{***}
Age 18-24	-0.49^{***}	-0.46^{***}	-0.35^{***}	-0.53^{***}
Age 25-34	-0.19^{***}	-0.17^{***}	-0.10^{***}	-0.22^{***}
Age 35-44	-0.04^{***}	-0.03^{***}	0.02	-0.07^{***}
# Children	0.003	-0.01	-0.05^{***}	0.02^{***}
# Children Under 5 Yrs	-0.03^{***}	-0.03^{***}	-0.02^{*}	-0.03^{**}
Ever Married	0.03^{***}	0.04^{***}	0.07^{***}	0.01
Employed Spouse	-0.11^{***}	-0.12^{***}	-0.18^{***}	-0.08^{***}
Veteran	-0.02^{***}	-0.02^{*}	0.01	-0.03^{*}
Single Mother	-0.17^{***}	-0.17^{***}	-0.14^{***}	-0.19^{***}
Asian	0.13^{***}	0.14^{***}	0.17^{***}	0.12^{***}
White	0.08^{***}	0.08^{***}	0.11^{***}	0.06^{***}
Black	-0.05^{***}	-0.05^{***}	-0.03^{*}	-0.06^{***}
Less than High School	-0.35^{***}	-0.35^{***}	-0.38^{***}	-0.33^{***}
Post-College Education	0.22^{***}	0.23^{***}	0.28^{***}	0.20^{***}
College-Educated Spouse	0.14^{***}	0.14^{***}	0.16^{***}	0.13^{***}
Veteran Spouse	-0.06^{***}	-0.05^{***}	-0.01	-0.08^{***}
Female X Age 18-24	0.02	0.02	0.05	0.01
Female X Age 25-34	0.03^{**}	0.03^{*}	0.06^{***}	0.02
Female X Age 35-44	0.04^{**}	0.04^{**}	0.05^{*}	0.03
Female X # Children	-0.04^{***}	-0.04^{***}	-0.04^{***}	-0.05^{***}
Female X # Children Under 5 Yrs	0.09^{***}	0.09^{***}	0.07^{***}	0.11^{***}
Female X Ever Married	-0.08^{***}	-0.08^{***}	-0.08^{***}	-0.07^{***}
Female X Employed Spouse	0.01	0.01	0.03^{*}	-0.01
Constant	3.05^{***}	2.96^{***}	2.51^{***}	3.33^{***}
Year Fixed effects	*Yes*	*Yes*	*Yes*	*Yes*
State Fixed effects	*No*	*No*	*No*	*No*
Observations	52,151	507,255	213,107	294,148
Adjusted R^2	0.21			
ρ		0.07	0.38	-0.15
Inverse Mills Ratio		0.03^{*}(0.01)	0.21^{***}(0.02)	-0.08^{***}(0.02)
Note:	$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.005$			

Discussion

- My selection model results largely confirm the literature:
  - probability of moving declines as workers age
  - more educated are more likely to move
  - migration is a family decision: spouse characteristics and number of children influence the decision to move
- Comparing my Uncorrected and Corrected models shows that while there is sample bias from migration selection, it has little practical effect on the wage equation of in-migrants.
- After correcting for sample bias, the wage equation of in-migrants has changed only slightly between Pre- and Post-COVID.

Discussion: Weaknesses

- ACS data cannot track people over time, so I cannot measure the effect of in-migration at the person level.
- ACS data only reports whether the respondent moved in the past year. I cannot identify how many times the respondent moved states, or how long they have been in the current state.
- The wage regressions are restricted to employed people and do not correct for labor force participation selection bias.
- I have not analyzed geography-specific moving decisions, e.g. accounting for moving costs or economic characteristics of the origin location.
- The wage models are limited to in-migrants only and cannot estimate the wage effects of moving states.

Conclusion

- ACS data can and do confirm existing literature results on the decision to migrate.
- My results show that after incorporating my control variables, selection does not substantially bias the wage equation of U.S. in-migrants.
- COVID does not appear to have substantially changed the wage equation of in-migrants.
- Future work should account for selection bias due to labor force non-participation, and should analyze origin-destination specific migration effects.
- Policy recommendation: cities/states seeking to increase their revenues should create tax incentives for young educated remote workers, with potentially additional benefits for married couples. These groups are more likely to move states.

References

- Borjas, George J., *Labor Economics*, McGraw Hill 9ed (2024).
- Borjas, George J., "Self-Selection and the Earnings of Immigrants," American Economic Review 77 (September 1987): 531-553.
- Heckman, James J., "Sample Selection Bias as a Specification Error," Econometrica 47 (January 1979): 153-61.
- Roy, Andrew D., "Some Thoughts on the Distribution of Earnings," Oxford Economic Papers 3 (June 1951): 135-146.

Appendix: Code

The code for this assignment is available here:
https://github.com/svanomm/labor-week-4-assignment.