Data Skills Evaluation Report

Xiling (Celia) Zhu[*]

November, 2020

1   Summary Statistics

Table 1 reports the counts of moves in the sample across US regions and between urban and non urban. From 1979 to 2012, $7,489$ respondents in the sample moved across regions, and $12,359$ respondents moved between urban and non-urban areas.

Table 2 reports the mean wage, employment, and education attainment in each region. In the period of 1979 to 2012, the average wage was the highest in Region 1 (Northeast) with 17581.75^{1}, and lowest in Region 3 (South) with 14436.54. In the four regions, the employment rates were around 80% to 83%, and average years of schooling were around 10 years, which were similar across all four regions. Region 2 (North Central) had the highest employment rate in the period of 1979 to 2012, and Region 1 (Northeast) had the highest education attainment.

Table 3 reports the mean wages, employment, and education attainment in urban and non-urban areas. From 1979 to 2012, urban areas had higher average wage, employment rate, and education attainment than non-urban areas.

---

[*]xiling@uchicago.edu

[1]The unit of wage is defined by this task, though the original NLSY 79 used actual dollars reported by respondents.

Table 1: Counts of Moves

	Count	Observations
Move across regions	7,489	228,667
Move between urban and non-urban areas	12,359	218,477

Table 2: Average Wage, Employment, and Education Attainment in Four Main Regions

	Wage	Employment	Education attainment
Region 1			
mean	17581.75	0.80	10.65
sd	75720.42	0.40	2.03
count	38,916	38,916	38,916
Region 2			
mean	15675.49	0.83	10.54
sd	49312.40	0.38	1.95
count	50,176	50,176	50,176
Region 3			
mean	14436.54	0.81	10.41
sd	44661.67	0.39	2.02
count	83,371	83,371	83,371
Region 4			
mean	16511.66	0.80	10.46
sd	69283.08	0.40	2.07
count	43,033	43,033	43,033

Table 3: Average Wage, Employment, and Education Attainment in Urban and Non Urban

	Wage	Employment	Education attainment
Non urban			
mean	14169.39	0.81	10.26
sd	34178.42	0.39	1.99
count	43,762	43,762	43,762
Urban			
mean	16722.4127	0.8319	10.5176
sd	63559.8966	0.3739	2.0321
count	164,664	164,664	164,664

2 Event Studies

Figure 1 reports the trend of wages for the respondents who moved across regions, in the five-year window centered at the moving year. Figure 2 reports the trend of wages for the respondents who moved between urban and non urban, in the five year window centered at the moving year.

To visually assess the effect of moving on wages, I only used the sub-sample of respondents who made exactly one move in the five-year window. Arguably, there would be selection bias if we include people who moved more frequently. For example, their frequent moves were probably attributable to the financial stress.

Generally, the two figures exhibit similar trends of wage in the five year window. On average, movers experienced a fall in wage in the first year of moving ($t = 0$), then their wage would gradually increase in the following two years. In the third year of their moving ($t = 2$), the average income surpassed the highest average income in the two years leading to the moving ($t = -2, -1$).

3

Except for other moving reasons such as college and military service, it's reasonable to assume movers in the sample were attracted by a higher wage. Then a question naturally arises: why did movers' average wage decrease in the first year of moving $(t = 0)$?

One hypothesis is that the wage decrease at $t = 0$ is caused by the loss of network, which might hurt job searching. Consequently, the movers had to accept a lower wage in their first year. The wage decrease for across-region movers was more drastic than that for urban-non urban movers. This is probably because the loss of social network was more significant when the respondent moved to a different region, while the moves between urban and non urban could be of shorter distance, and the movers were still able to maintain part of his/her network.

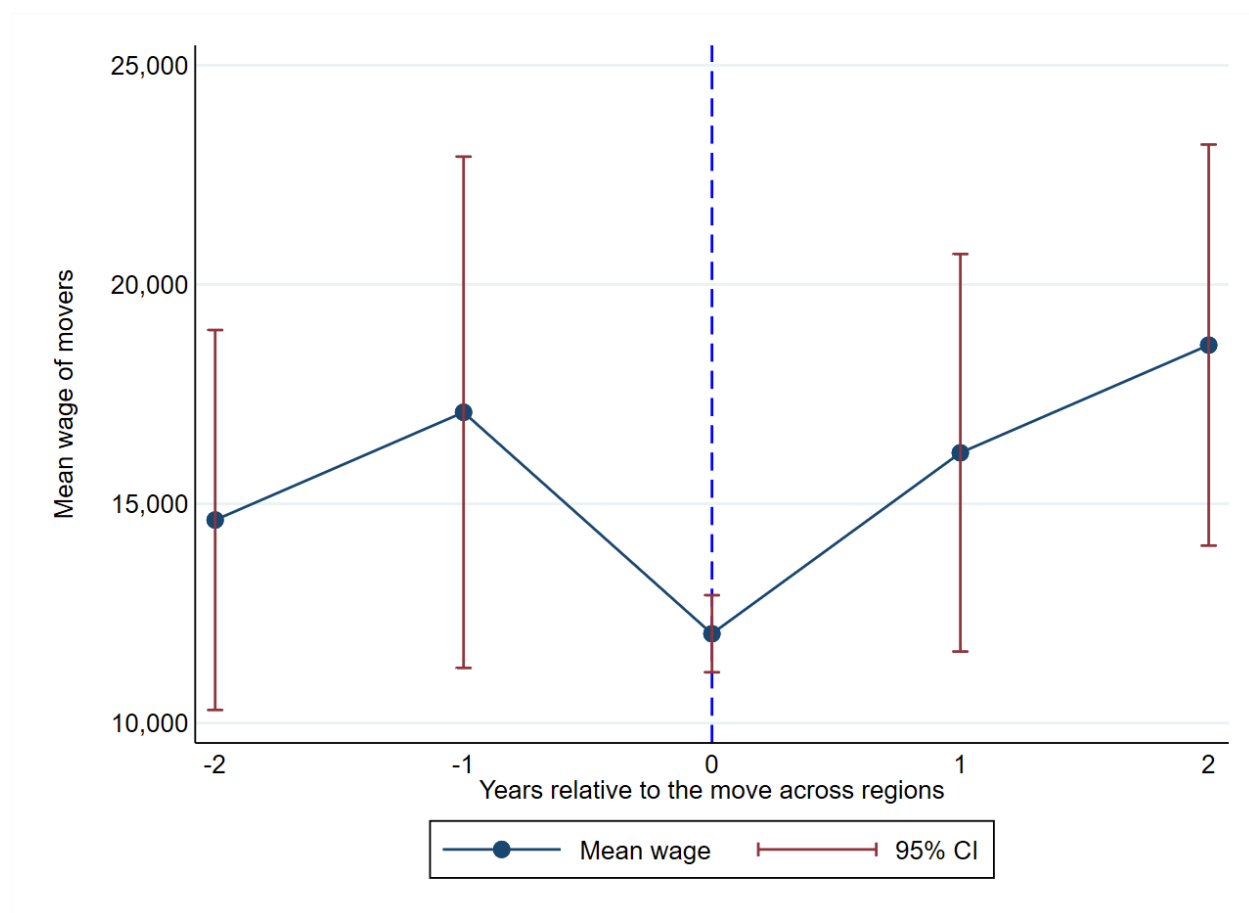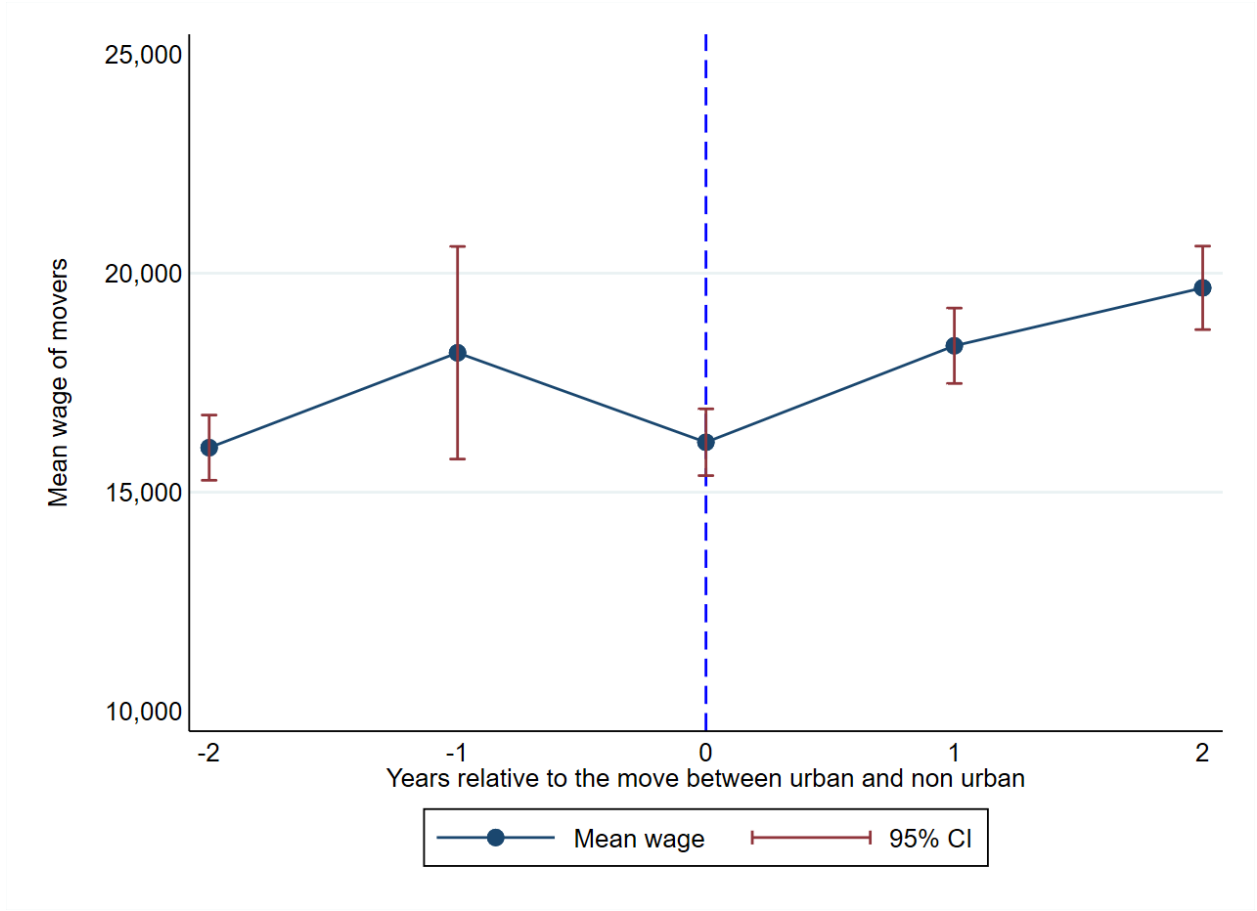Figure 1: Average wage of movers across regions

Figure 2: Average wage of movers between urban and non urban

3 Comparing Movers to Stayers

Since I had no information on the sampling process of NLSY 79, which was a survey data, I first assumed it is a random sample. Table 4 and 6 reports the differences between the wage of movers and stayers in both region or origin (table 4) and region of arrival (table 6), free of clusters. Table 5 and 7 reports the differences between the wage of movers and stayers in both region or origin (table 5) and region of arrival (table 7), assuming wage residual was clustered on cohort.

The model I used for comparison is adapted from the Mincer equation.

To compare movers and stayers in the region of origin:

$$lnw = lnw_0 + \alpha Movefrom_r + \beta_1 educ + \beta_2 experience + \beta_3 experience^2 + \beta_4 gender + \epsilon \quad (1)$$

To compare movers and stayers in the region of arrival:

$$lnw = lnw_0 + \alpha Moveto_r + \beta_1 educ + \beta_2 experience + \beta_3 experience^2 + \beta_4 gender + \epsilon \quad (2)$$

Where $Movefrom_r$ is a dummy variable indicating the movers and stayers: $Movefrom_r = 1$ if the respondent moved out of region r, $Movefrom_r = 0$ if the respondent stayed in the of region r. Similarly, $Moveto_r = 1$ if the respondent moved into region r, $Moveto_r = 0$ if the respondent stayed in the of region r.

Age was used as a proxy for experience.

The rate of return from both moving out of and into region 1 is negative. The rate of return from moving out of region 4 is negative, though the effect was not significant when clustered on cohort.

Table 4: Compare movers out of a region to stayers at the same region (no cluster of wage residulas)

	Log wage	Log wage	Log wage	Log wage
Move from region 1	-0.162***			
	(0.0123)			
Move from region 2		0.00948		
		(0.0119)		
Move from region 3			0.00801	
			(0.00966)	
Move from region 4				-0.0785***
				(0.0130)
Age	0.398***	0.361***	0.371***	0.346***
	(0.00510)	(0.00474)	(0.00363)	(0.00512)
Age squared	-0.00465***	-0.00414***	-0.00432***	-0.00395***
	(0.0000733)	(0.0000683)	(0.0000526)	(0.0000739)
Education attainment	0.0402***	0.0471***	0.0501***	0.0344***
	(0.00295)	(0.00281)	(0.00208)	(0.00287)
Gender	-0.420***	-0.574***	-0.422***	-0.525***
	(0.0117)	(0.0108)	(0.00824)	(0.0118)
Constant	2.085***	2.668***	2.255***	3.040***
	(0.0879)	(0.0826)	(0.0628)	(0.0888)
Observations	33452	42146	67319	34674

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Compare movers out of a region to stayers at the same region (cluster wage residuals on cohort)

	Log wage	Log wage	Log wage	Log wage
Move from region 1	-0.162***			
	(0.0212)			
Move from region 2		0.00948		
		(0.0290)		
Move from region 3			0.00801	
			(0.0157)	
Move from region 4				-0.0785
				(0.0343)
Age	0.398***	0.361***	0.371***	0.346***
	(0.0255)	(0.0299)	(0.0223)	(0.0232)
Age squared	-0.00465***	-0.00414***	-0.00432***	-0.00395***
	(0.000404)	(0.000440)	(0.000332)	(0.000362)
Education attainment	0.0402	0.0471	0.0501	0.0344
	(0.0246)	(0.0218)	(0.0215)	(0.0192)
Gender	-0.420***	-0.574***	-0.422***	-0.525***
	(0.0436)	(0.0566)	(0.0186)	(0.0315)
Constant	2.085***	2.668***	2.255***	3.040***
	(0.383)	(0.488)	(0.317)	(0.296)
Observations	33452	42146	67319	34674

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Compare movers into a region to stayers at the same region (no cluster of wage residuals

	Log wage	Log wage	Log wage	Log wage
Move to region 1	-0.201***			
	(0.0134)			
Move to region 2		-0.0194		
		(0.0127)		
Move to region 3			0.0142	
			(0.00914)	
Move to region 4				-0.0223
				(0.0124)
Age	0.404***	0.361***	0.369***	0.347***
	(0.00536)	(0.00482)	(0.00355)	(0.00502)
Age squared	-0.00473***	-0.00413***	-0.00430***	-0.00396***
	(0.0000771)	(0.0000695)	(0.0000514)	(0.0000724)
Education attainment	0.0378***	0.0456***	0.0490***	0.0397***
	(0.00310)	(0.00286)	(0.00204)	(0.00280)
Gender	-0.408***	-0.574***	-0.426***	-0.533***
	(0.0123)	(0.0110)	(0.00807)	(0.0115)
Constant	1.992***	2.671***	2.319***	2.982***
	(0.0923)	(0.0844)	(0.0614)	(0.0869)
Observations	31486	39860	69861	36389

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Compare movers into a region to stayers at the same region (cluster wage residuals on cohort)

	Log wage	Log wage	Log wage	Log wage
Move to region 1	-0.201***			
	(0.0286)			
Move to region 2		-0.0194		
		(0.0202)		
Move to region 3			0.0142	
			(0.0131)	
Move to region 4				-0.0223
				(0.0362)
Age	0.404***	0.361***	0.369***	0.347***
	(0.0245)	(0.0289)	(0.0236)	(0.0227)
Age squared	-0.00473***	-0.00413***	-0.00430***	-0.00396***
	(0.000390)	(0.000428)	(0.000350)	(0.000355)
Education attainment	0.0378	0.0456	0.0490	0.0397
	(0.0266)	(0.0223)	(0.0217)	(0.0181)
Gender	-0.408***	-0.574***	-0.426***	-0.533***
	(0.0529)	(0.0572)	(0.0147)	(0.0297)
Constant	1.992**	2.671***	2.319***	2.982***
	(0.397)	(0.493)	(0.337)	(0.268)
Observations	31486	39860	69861	36389

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$