BFI Predoc Research | Evaluation Task Report
Migration and Wages Across US Regions

Lutfi Sun

November 7, 2020

1 Summary Statistics

1.1 Moves Across US Regions

Table 1: Migration Between Regions

		Origin				
		Region 1	Region 2	Region 3	Region 4	All (+NA)
Count	Destination					
	Region 1	1743	252	947	380	3373
	Region 2	305	3269	859	478	4968
	Region 3	1227	1024	7163	953	10484
	Region 4	428	553	927	2519	4468
	Total	3703	5098	9896	4330	23293
Percent	Destination					
	Region 1	47.07	4.94	9.57	8.78	14.48
	Region 2	8.24	64.12	8.68	11.04	21.33
	Region 3	33.14	20.09	72.38	22.01	45.01
	Region 4	11.56	10.85	9.37	58.18	19.18
	Total	100	100	100	100	100

Most frequently, we observe within-region migration. I was able to calculate this by looking subsequent observations for an individual where urban type changes but region stays the same. This probably is an understatement of within-region moves, however, as there might be moves where neither region nor urban type change.

Out of the four regions, region 3 received the highest count of move-ins, 10484, which makes up about 45 percent of all move-ins. Region 3 also experienced highest number of move-outs, but most of those (72.4 percent) were within-region migration.

The number of moves from Region 1 exceeds the number of moves to it suggesting a negative net migration. Region 3 and Region 4, on the other hand, experience a positive net migration. There seems to be a strong link between Region 1 and Region 3 where many people (possibly same individuals) move between them.

1.2 Moves Across Urban and Non-Urban Areas

Table 2: Migration Between Urban and Rural Areas

	Origin			
	Urban 0	Urban 1	Urban 2	All
Destination				
Urban 0	308	6507	385	7514
Urban 1	6268	3982	792	12422
Urban 2	556	854	3	1425
Total	7132	11343	1180	21361

In Table 2, we see that Urban 1 areas receive the highest number of move-ins whereas Urban 2 areas receive the lowest.

We see urban-rural migration is frequent both ways:

- Rural to Urban: 6268 + 556 = 6824 moves.
- Urban to Rural: 6507 + 385 = 6892 moves.

Rural to Rural is the rarest move.

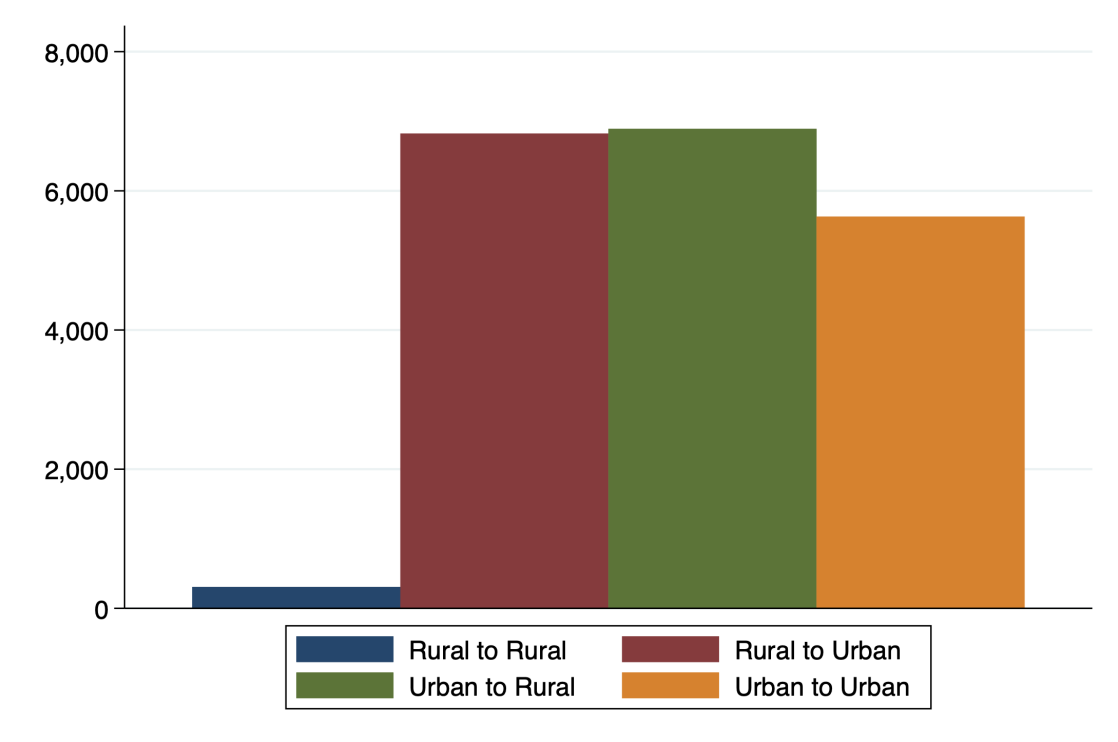


Figure 1: Migration Between Urban and Rural Areas

1.3 Socioeconomic Variables Between Regions

Table 3: Socioeconomic Variables Across Regions

		Mean	(sd)			Mean	(sd)
Region 1				Region 3			
	Wage	17581.75	(75720)		Wage	14436.54	(44661)
	Education	10.65	(2.03)		Education	10.41	(2.02)
	Employment	79.78	(40.16)		Employment	80.93	(39.29)
Region 2				Region 4			
	Wage	15675.49	(49312)		Wage	16511.66	(69283)
	Education	10.54	(1.95)		Education	10.46	(2.07)
	Employment	82.57	(37.94)		Employment	79.70	(40.23)
Obs		215496				215496	

We observe the highest mean wage in Region 1 where it is above $17k a year. Region 1 also has the most educated population and the lowest employment rate. Perhaps low employment rates and high prices are among reasons why people migrate away from Region 1.

Region 2 is the most educated after Region 1 and has the highest employment rate among the four regions. The lowest mean wage is in Region 3 with around $14k. Perhaps many people move here not necessarily for higher wages but for higher chances of being employed.

Table 4: Socioeconomic Variables in Urban vs Rural Areas

		Mean	(sd)			Mean	(sd)
Urban 0				Urban 2			
	Wage	14170	(34178)		Wage	21192	(20566)
	Education	10.26	(1.99)		Education	10.03	(1.86)
	Employment	81.42	(38.89)		Employment	79.61	(40.39)
Urban 1				Total			
	Wage	16722	(63560)		Wage	16191	(58609)
	Education	10.52	(2.03)		Education	10.46	(2.03)
	Employment	83.19	(37.39)		Employment	82.82	(37.72)
Observations		208632				208632	

On average, observe higher wages as urban measure increases. Rural areas have the lowest mean wage with about $14k a year. Most urban areas (Urban 2) have the highest mean wages and lowest employment rates. Urban 1 areas have the highest educational attainment and highest employment rate levels.

2 Event Studies

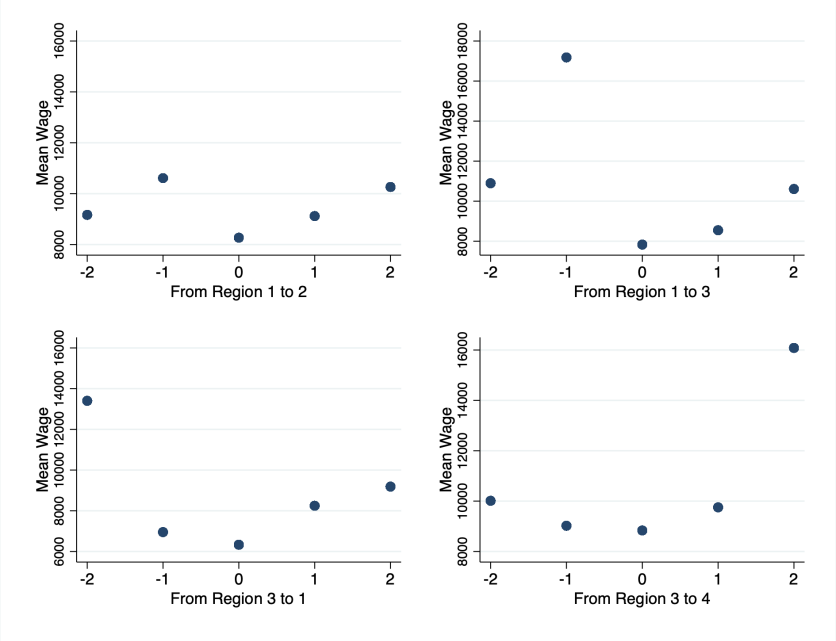


Figure 2: How Wages Change Before and After Move Across Regions

Wages on the year of move fall rapidly but starts rising the following year. It seems that moving from Region 3 to Region 1 saved some people from a decreasing trend in wages. The movers from Region 3 to Region 4 experienced a rapid rise in their wages.

Those who moved between rural areas (Urban = 0) had their wages decreasing on average before the move. After the move, the wages stabilized. Urban-Urban move on the other hand seemed to have no impact except the fall in moving year.

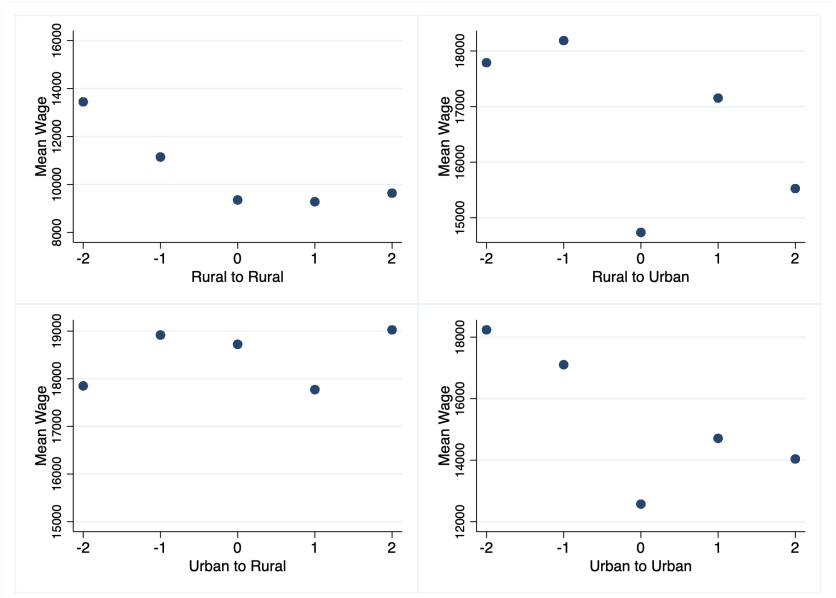


Figure 3: How Wages Change Before and After Move Between Urban and Rural

3 Comparing Movers to Stayers

Table 5: The Effect of Moving on Wages In Comparison with Stayers at Origin

	Origin Region 1 ln(wage)	Origin Region 2 ln(wage)	Origin Region 3 ln(wage)	Origin Region 4 ln(wage)
Age	0.388***	0.388***	0.388***	0.388***
	(0.00)	(0.00)	(0.00)	(0.00)
Age Squared	-0.005***	-0.005***	-0.005***	-0.005***
	(0.00)	(0.00)	(0.00)	(0.00)
Education Sqrd	0.002***	0.002***	0.002***	0.002***
	(0.00)	(0.00)	(0.00)	(0.00)
Gender	-0.455***	-0.455***	-0.455***	-0.455***
	(0.01)	(0.01)	(0.01)	(0.01)
Mover at 0	-0.089***	-0.090***	-0.081***	-0.071***
	(0.01)	(0.01)	(0.01)	(0.01)
Mover at 1	0.019	0.085**	0.004	-0.060
	(0.04)	(0.03)	(0.03)	(0.04)
Mover at 2	0.012	0.109***	0.027	-0.060
	(0.03)	(0.03)	(0.02)	(0.03)
Mover at 3	0.103***	0.117***	0.033	-0.064*
	(0.03)	(0.03)	(0.02)	(0.03)
Mover at 4	0.116***	0.073**	0.056***	0.011
	(0.03)	(0.02)	(0.02)	(0.03)
Regional FE	Yes	Yes	Yes	Yes
Constant	2.495***	2.505***	2.499***	2.499***
	(0.04)	(0.04)	(0.04)	(0.04)
R-sqr	0.369	0.369	0.369	0.369

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

.

We again observe a stark effect on the year of the move. The effect disappears or decreases a year after the move and seems to die down following the second year.

Those who moved from Region 2 observed higher wages right after the first year relative to those who stayed. Those who moved from Region 1 and Region 3 started having higher wages three to four years after the move.

Region 4, on the other hand, seems to be an anomaly where movers from that origin have experienced a more persistent negative effect on their wages.

In Table 6, we see a similar picture where the movers had between 20 to 40 percent drop in their wages. The highest falls are followed by highest gains: 40 percent fall followed by 34 percent gain relative to those who were native in Region 1. Similarly those who moved to region four were doing 38 percent better than those who were there a year after move in.

Table 6: The Effect of Moving on Wages In Comparison with Stayers at Destination

	Arrival Region 1 ln_wage	Arrival Region 2 ln_wage	Arrival Region 3 ln_wage	Arrival Region 4 ln_wage
Age	0.385***	0.385***	0.385***	0.385***
	(0.00)	(0.00)	(0.00)	(0.00)
Age Squared	-0.004***	-0.004***	-0.004***	-0.004***
	(0.00)	(0.00)	(0.00)	(0.00)
Education Sqrd	0.002***	0.002***	0.002***	0.002***
	(0.00)	(0.00)	(0.00)	(0.00)
Gender	-0.453***	-0.454***	-0.454***	-0.454***
	(0.01)	(0.01)	(0.01)	(0.01)
Mover at 0	-0.401***	-0.242***	-0.280***	-0.399***
	(0.03)	(0.03)	(0.02)	(0.03)
Mover at 1	0.343***	0.156***	0.264***	0.379***
	(0.04)	(0.04)	(0.03)	(0.04)
Mover at 2	0.042	-0.031	-0.005	0.002
	(0.04)	(0.03)	(0.03)	(0.03)
Mover at 3	-0.053	0.059	0.004	-0.045
	(0.03)	(0.03)	(0.02)	(0.03)
Mover at 4	0.027	0.035	-0.038*	-0.057*
	(0.03)	(0.02)	(0.02)	(0.02)
Regional FE	Yes	Yes	Yes	Yes
Constant	2.547***	2.517***	2.524***	2.520***
	(0.04)	(0.04)	(0.04)	(0.04)
R-sqr	0.369	0.368	0.369	0.369

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

4 Appendix

Table 7: Wage Across Regions

	(1) Wage b/se	(2) Employment pct b/se	(3) Education b/se
Region 2	-1906.35***	2.74***	-0.10***
	(442.16)	(0.24)	(0.01)
Region 3	-3138.21***	1.62***	-0.23***
	(413.53)	(0.22)	(0.01)
Region 4	-1077.93*	0.12	-0.17***
	(508.33)	(0.25)	(0.01)
Constant (Reg 1)	17574.23***	81.81***	10.53***
	(383.52)	(0.19)	(0.01)

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

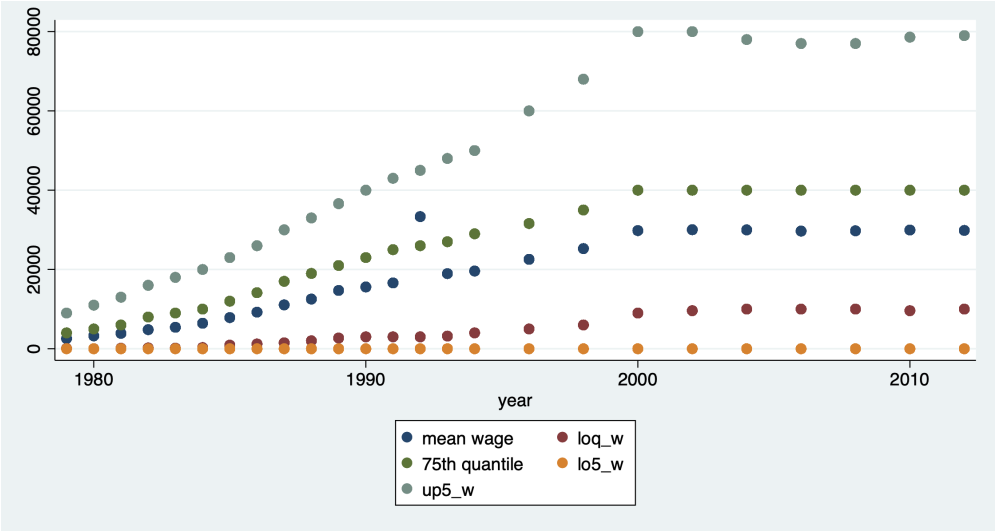
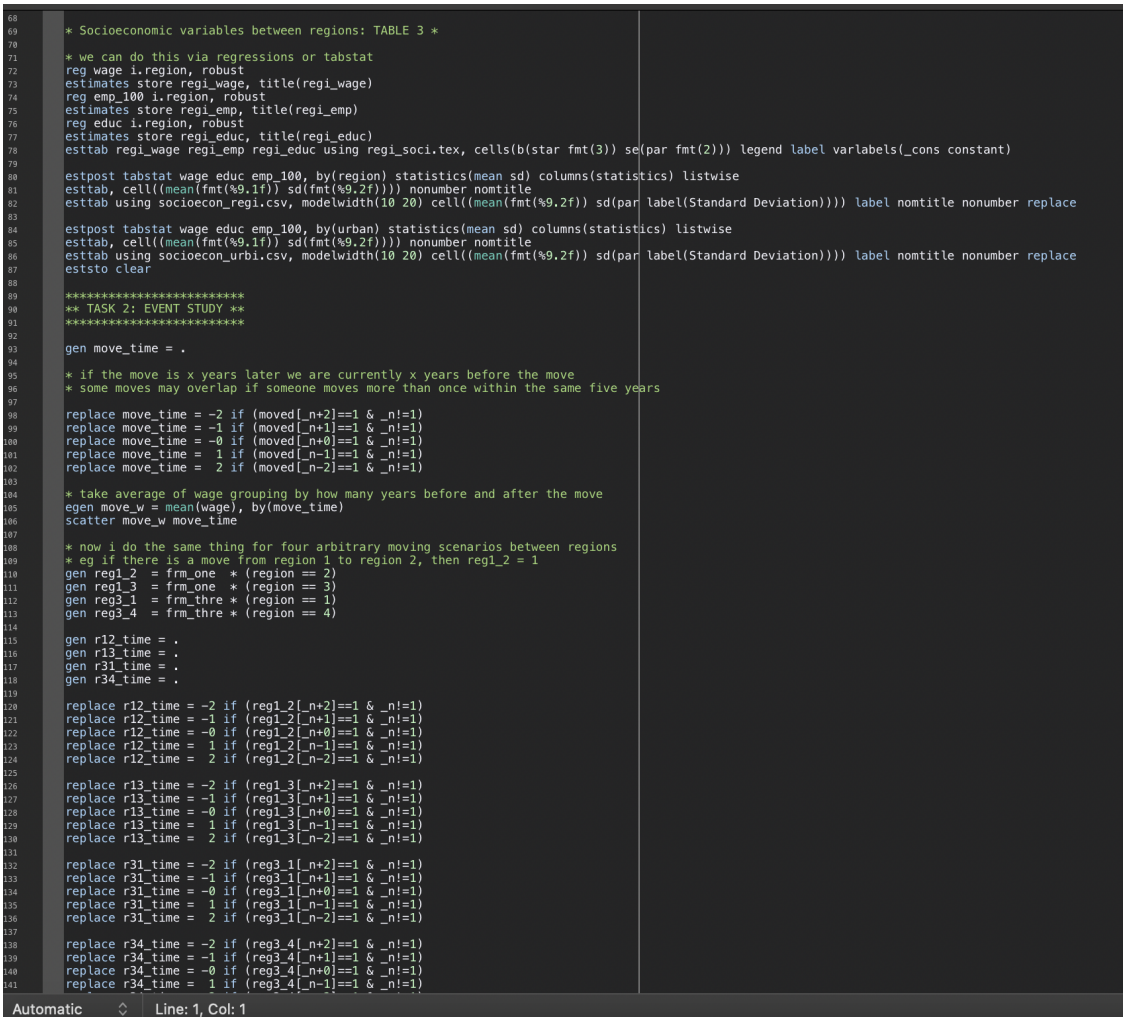


Figure 4: Wages are Stagnant and Wage Gap Widens

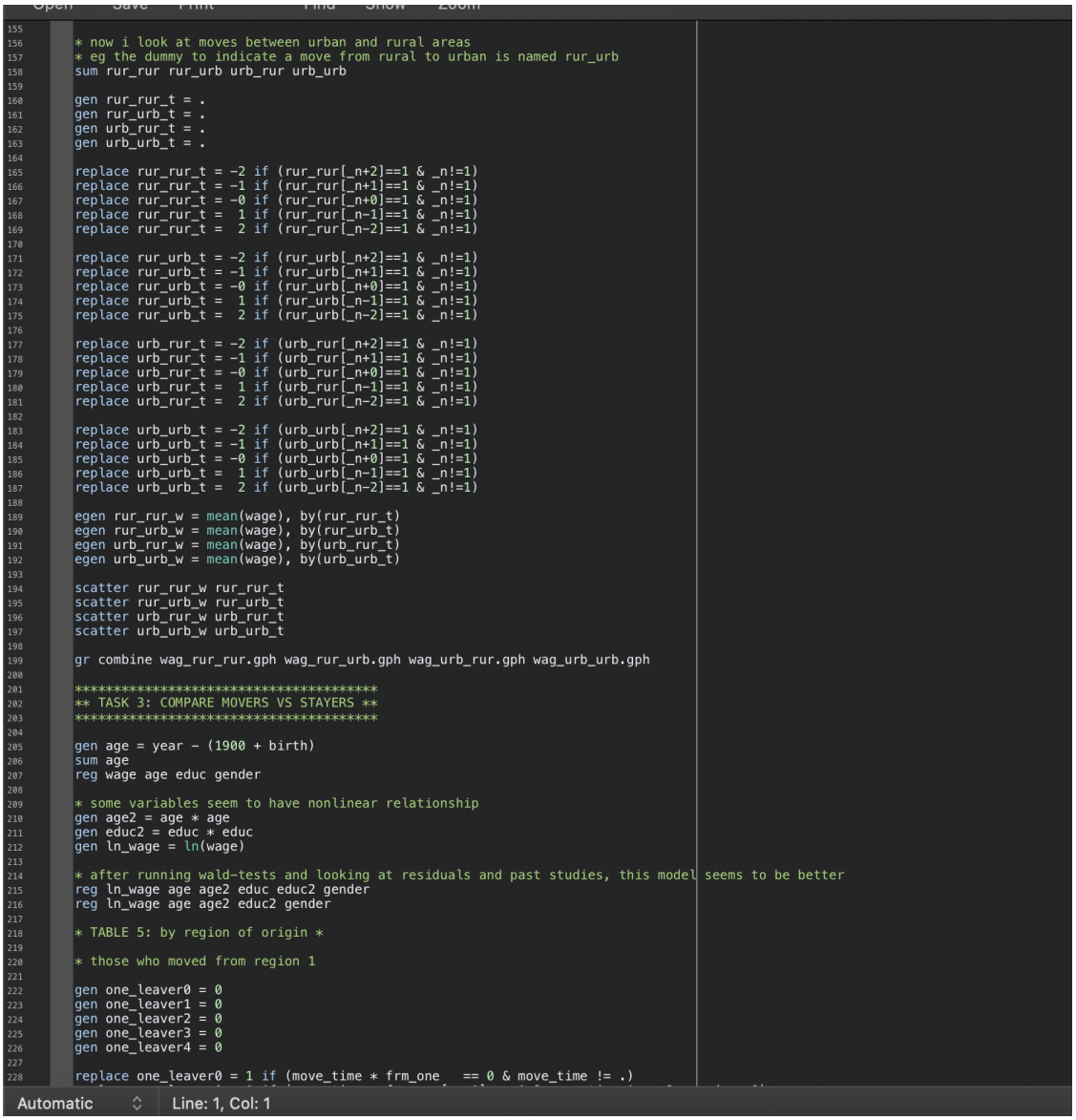
```
* Lutfi Sun | Nov 7, 2020 | BFI Labor Data Task *

** HOUSE KEEPING **

clear
cd "/Users/lutfisun/Desktop/BFI Labor Data Task"

ssc install estout
eststo clear

* numericcols option will replace NA entries with missing values that stata can understand
import delimited "nlsy79-prepared.csv", numericcols(5 6 7 8 9)

* focus on those we have information about the move
drop if missing(region) & missing(urban)
* Define employment as wage>0
gen employed = (wage > 0)
gen emp_100 = employed * 100

*********************************
** TASK 1: SUMMARY STATISTICS **
*********************************

* changes in region or urban for the same individual would imply moving
* in this data, we cannot observe moves within the same region and to same type of area (eg rural to rural)
* using lag and logical test to see when and whether an individual moved
bysort i (year): gen moved = (region-region[_n-1] != 0) | (urban-urban[_n-1] != 0 & !missing(urban)) if _n!=1

* see if one moves to a more or less urban area
gen urb_move = moved * (urban-urban[_n-1])

* what urban area the individual left
gen urb0_x = moved * (urban[_n-1] == 0)
gen urb1_x = moved * (urban[_n-1] == 1)
gen urb2_x = moved * (urban[_n-1] == 2)

* if we want to look deeper into urban, we can generate a variable for each of the nine ways to move between rural and urban regions
gen urb0_0 = moved * (urban-urban[_n-1] ==  0) * (urban == 0)
gen urb0_1 = moved * (urban-urban[_n-1] ==  1) * (urban == 1)
gen urb0_2 = moved * (urban-urban[_n-1] ==  2) * (urban == 2)

gen urb1_0 = moved * (urban-urban[_n-1] == -1) * (urban == 0)
gen urb1_1 = moved * (urban-urban[_n-1] ==  0) * (urban == 1)
gen urb1_2 = moved * (urban-urban[_n-1] ==  1) * (urban == 2)

gen urb2_0 = moved * (urban-urban[_n-1] == -2) * (urban == 0)
gen urb2_1 = moved * (urban-urban[_n-1] == -1) * (urban == 1)
gen urb2_2 = moved * (urban-urban[_n-1] ==  0) * (urban == 2)

gen rur_rur = urb0_0
gen rur_urb = urb0_1 + urb0_2
gen urb_rur = urb1_0 + urb2_0
gen urb_urb = urb1_1 + urb1_2 + urb2_1 + urb2_2

* what region the individual left
gen frm_one  = moved * (region[_n-1] == 1)
gen frm_two  = moved * (region[_n-1] == 2)
gen frm_thre = moved * (region[_n-1] == 3)
gen frm_four = moved * (region[_n-1] == 4)

* Migration between regions: TABLE 1 *
estpost tabstat frm_one frm_two frm_thre frm_four moved, by(region) stat(sum)

graph bar (sum) frm_one frm_two frm_thre frm_four, over(region)

* Migration between area types: TABLE 2 *
estpost tabstat urb0_x urb1_x urb2_x moved, by(urban) stat(sum)

* Socioeconomic variables between regions: TABLE 3 *

* we can do this via regressions or tabstat
reg wage i.region, robust
estimates store regi_wage, title(regi_wage)
reg emp_100 i.region, robust
```

Figure 5: Do File Also Included in the Upload



```
* Socioeconomic variables between regions: TABLE 3 *

* we can do this via regressions or tabstat
reg wage i.region, robust
estimates store regi_wage, title(regi_wage)
reg emp_100 i.region, robust
estimates store regi_emp, title(regi_emp)
reg educ i.region, robust
estimates store regi_educ, title(regi_educ)
esttab regi_wage regi_emp regi_educ using regi_soci.tex, cells(b(star fmt(3)) se(par fmt(2))) legend label varlabels(_cons constant)

estpost tabstat wage_educ emp_100, by(region) statistics(mean sd) columns(statistics) listwise
esttab, cell((mean(fmt(%9.1f)) sd(fmt(%9.2f)))) nonumber nomtitle
esttab using socioecon_regi.csv, modelwidth(10 20) cell((mean(fmt(%9.2f)) sd(par label(Standard Deviation)))) label nomtitle nonumber replace

estpost tabstat wage_educ emp_100, by(urban) statistics(mean sd) columns(statistics) listwise
esttab, cell((mean(fmt(%9.1f)) sd(fmt(%9.2f)))) nonumber nomtitle
esttab using socioecon_urbi.csv, modelwidth(10 20) cell((mean(fmt(%9.2f)) sd(par label(Standard Deviation)))) label nomtitle nonumber replace
eststo clear

************************
** TASK 2: EVENT STUDY **
************************

gen move_time = .

* if the move is x years later we are currently x years before the move
* some moves may overlap if someone moves more than once within the same five years

replace move_time = -2 if (moved[_n+2]==1 & _n!=1)
replace move_time = -1 if (moved[_n+1]==1 & _n!=1)
replace move_time = -0 if (moved[_n+0]==1 & _n!=1)
replace move_time =  1 if (moved[_n-1]==1 & _n!=1)
replace move_time =  2 if (moved[_n-2]==1 & _n!=1)

* take average of wage grouping by how many years before and after the move
egen move_w = mean(wage), by(move_time)
scatter move_w move_time

* now i do the same thing for four arbitrary moving scenarios between regions
* eg if there is a move from region 1 to region 2, then reg1_2 = 1
gen reg1_2  = frm_one  * (region == 2)
gen reg1_3  = frm_one  * (region == 3)
gen reg3_1  = frm_thre * (region == 1)
gen reg3_4  = frm_thre * (region == 4)

gen r12_time = .
gen r13_time = .
gen r31_time = .
gen r34_time = .

replace r12_time = -2 if (reg1_2[_n+2]==1 & _n!=1)
replace r12_time = -1 if (reg1_2[_n+1]==1 & _n!=1)
replace r12_time = -0 if (reg1_2[_n+0]==1 & _n!=1)
replace r12_time =  1 if (reg1_2[_n-1]==1 & _n!=1)
replace r12_time =  2 if (reg1_2[_n-2]==1 & _n!=1)

replace r13_time = -2 if (reg1_3[_n+2]==1 & _n!=1)
replace r13_time = -1 if (reg1_3[_n+1]==1 & _n!=1)
replace r13_time = -0 if (reg1_3[_n+0]==1 & _n!=1)
replace r13_time =  1 if (reg1_3[_n-1]==1 & _n!=1)
replace r13_time =  2 if (reg1_3[_n-2]==1 & _n!=1)

replace r31_time = -2 if (reg3_1[_n+2]==1 & _n!=1)
replace r31_time = -1 if (reg3_1[_n+1]==1 & _n!=1)
replace r31_time = -0 if (reg3_1[_n+0]==1 & _n!=1)
replace r31_time =  1 if (reg3_1[_n-1]==1 & _n!=1)
replace r31_time =  2 if (reg3_1[_n-2]==1 & _n!=1)

replace r34_time = -2 if (reg3_4[_n+2]==1 & _n!=1)
replace r34_time = -1 if (reg3_4[_n+1]==1 & _n!=1)
replace r34_time = -0 if (reg3_4[_n+0]==1 & _n!=1)
replace r34_time =  1 if (reg3_4[_n-1]==1 & _n!=1)
```

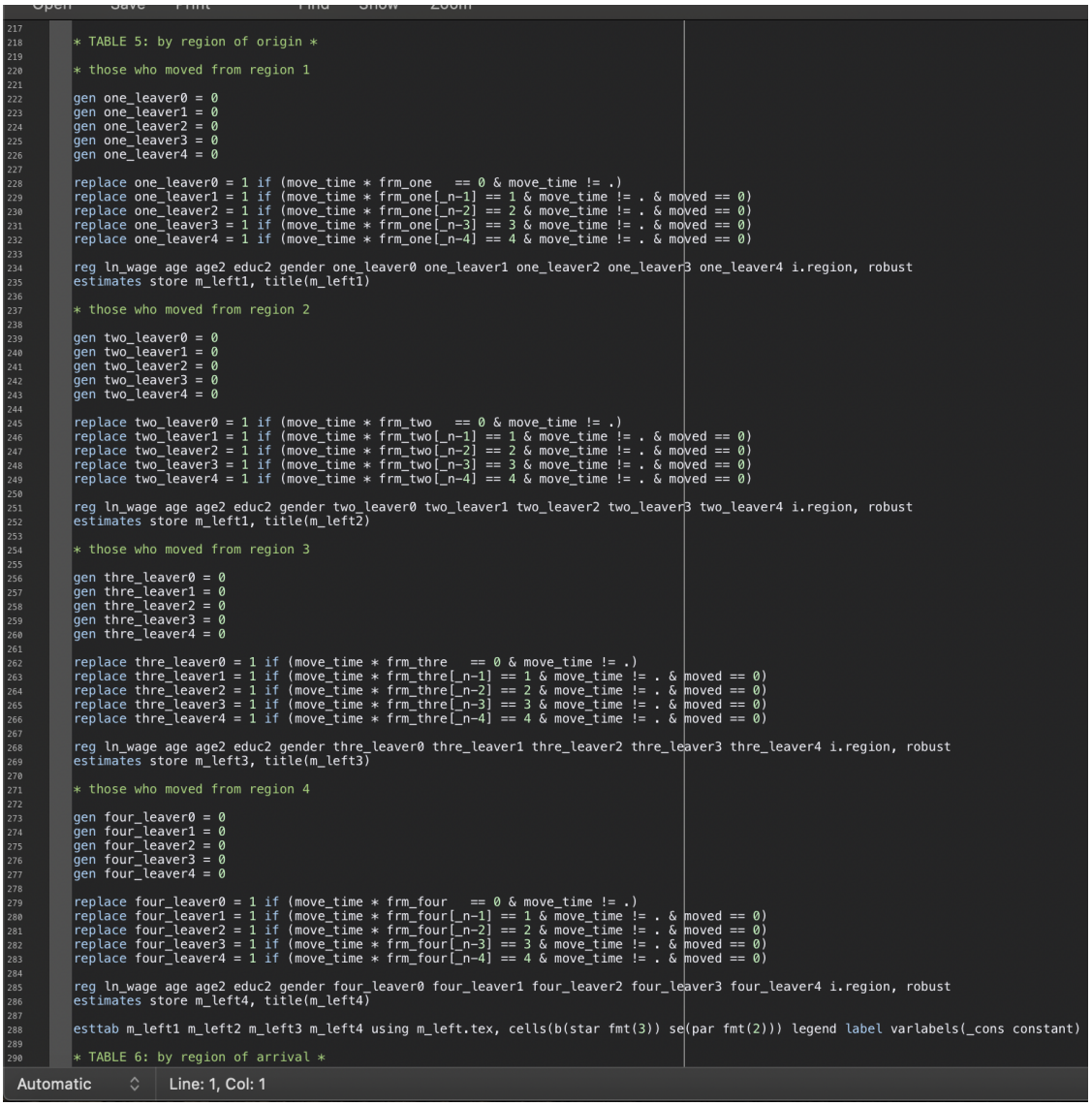

Figure 6: Do File Also Included in the Upload



```
* now i look at moves between urban and rural areas
* eg the dummy to indicate a move from rural to urban is named rur_urb
sum rur_rur rur_urb urb_rur urb_urb

gen rur_rur_t = .
gen rur_urb_t = .
gen urb_rur_t = .
gen urb_urb_t = .

replace rur_rur_t = -2 if (rur_rur[_n+2]==1 & _n!=1)
replace rur_rur_t = -1 if (rur_rur[_n+1]==1 & _n!=1)
replace rur_rur_t = -0 if (rur_rur[_n+0]==1 & _n!=1)
replace rur_rur_t =  1 if (rur_rur[_n-1]==1 & _n!=1)
replace rur_rur_t =  2 if (rur_rur[_n-2]==1 & _n!=1)

replace rur_urb_t = -2 if (rur_urb[_n+2]==1 & _n!=1)
replace rur_urb_t = -1 if (rur_urb[_n+1]==1 & _n!=1)
replace rur_urb_t = -0 if (rur_urb[_n+0]==1 & _n!=1)
replace rur_urb_t =  1 if (rur_urb[_n-1]==1 & _n!=1)
replace rur_urb_t =  2 if (rur_urb[_n-2]==1 & _n!=1)

replace urb_rur_t = -2 if (urb_rur[_n+2]==1 & _n!=1)
replace urb_rur_t = -1 if (urb_rur[_n+1]==1 & _n!=1)
replace urb_rur_t = -0 if (urb_rur[_n+0]==1 & _n!=1)
replace urb_rur_t =  1 if (urb_rur[_n-1]==1 & _n!=1)
replace urb_rur_t =  2 if (urb_rur[_n-2]==1 & _n!=1)

replace urb_urb_t = -2 if (urb_urb[_n+2]==1 & _n!=1)
replace urb_urb_t = -1 if (urb_urb[_n+1]==1 & _n!=1)
replace urb_urb_t = -0 if (urb_urb[_n+0]==1 & _n!=1)
replace urb_urb_t =  1 if (urb_urb[_n-1]==1 & _n!=1)
replace urb_urb_t =  2 if (urb_urb[_n-2]==1 & _n!=1)

egen rur_rur_w = mean(wage), by(rur_rur_t)
egen rur_urb_w = mean(wage), by(rur_urb_t)
egen urb_rur_w = mean(wage), by(urb_rur_t)
egen urb_urb_w = mean(wage), by(urb_urb_t)

scatter rur_rur_w rur_rur_t
scatter rur_urb_w rur_urb_t
scatter urb_rur_w urb_rur_t
scatter urb_urb_w urb_urb_t

gr combine wag_rur_rur.gph wag_rur_urb.gph wag_urb_rur.gph wag_urb_urb.gph

****************************************
** TASK 3: COMPARE MOVERS VS STAYERS **
****************************************

gen age = year - (1900 + birth)
sum age
reg wage age educ gender

* some variables seem to have nonlinear relationship
gen age2 = age * age
gen educ2 = educ * educ
gen ln_wage = ln(wage)

* after running wald-tests and looking at residuals and past studies, this model seems to be better
reg ln_wage age age2 educ educ2 gender
reg ln_wage age age2 educ2 gender

* TABLE 5: by region of origin *

* those who moved from region 1

gen one_leaver0 = 0
gen one_leaver1 = 0
gen one_leaver2 = 0
gen one_leaver3 = 0
gen one_leaver4 = 0

replace one_leaver0 = 1 if (move_time * frm_one   == 0 & move_time != .)
```

Automatic Line: 1, Col: 1

Figure 7: Do File Also Included in the Upload



```
* TABLE 5: by region of origin *

* those who moved from region 1

gen one_leaver0 = 0
gen one_leaver1 = 0
gen one_leaver2 = 0
gen one_leaver3 = 0
gen one_leaver4 = 0

replace one_leaver0 = 1 if (move_time * frm_one   == 0 & move_time != .)
replace one_leaver1 = 1 if (move_time * frm_one[_n-1] == 1 & move_time != . & moved == 0)
replace one_leaver2 = 1 if (move_time * frm_one[_n-2] == 2 & move_time != . & moved == 0)
replace one_leaver3 = 1 if (move_time * frm_one[_n-3] == 3 & move_time != . & moved == 0)
replace one_leaver4 = 1 if (move_time * frm_one[_n-4] == 4 & move_time != . & moved == 0)

reg ln_wage age age2 educ2 gender one_leaver0 one_leaver1 one_leaver2 one_leaver3 one_leaver4 i.region, robust
estimates store m_left1, title(m_left1)

* those who moved from region 2

gen two_leaver0 = 0
gen two_leaver1 = 0
gen two_leaver2 = 0
gen two_leaver3 = 0
gen two_leaver4 = 0

replace two_leaver0 = 1 if (move_time * frm_two   == 0 & move_time != .)
replace two_leaver1 = 1 if (move_time * frm_two[_n-1] == 1 & move_time != . & moved == 0)
replace two_leaver2 = 1 if (move_time * frm_two[_n-2] == 2 & move_time != . & moved == 0)
replace two_leaver3 = 1 if (move_time * frm_two[_n-3] == 3 & move_time != . & moved == 0)
replace two_leaver4 = 1 if (move_time * frm_two[_n-4] == 4 & move_time != . & moved == 0)

reg ln_wage age age2 educ2 gender two_leaver0 two_leaver1 two_leaver2 two_leaver3 two_leaver4 i.region, robust
estimates store m_left1, title(m_left2)

* those who moved from region 3

gen thre_leaver0 = 0
gen thre_leaver1 = 0
gen thre_leaver2 = 0
gen thre_leaver3 = 0
gen thre_leaver4 = 0

replace thre_leaver0 = 1 if (move_time * frm_thre   == 0 & move_time != .)
replace thre_leaver1 = 1 if (move_time * frm_thre[_n-1] == 1 & move_time != . & moved == 0)
replace thre_leaver2 = 1 if (move_time * frm_thre[_n-2] == 2 & move_time != . & moved == 0)
replace thre_leaver3 = 1 if (move_time * frm_thre[_n-3] == 3 & move_time != . & moved == 0)
replace thre_leaver4 = 1 if (move_time * frm_thre[_n-4] == 4 & move_time != . & moved == 0)

reg ln_wage age age2 educ2 gender thre_leaver0 thre_leaver1 thre_leaver2 thre_leaver3 thre_leaver4 i.region, robust
estimates store m_left3, title(m_left3)

* those who moved from region 4

gen four_leaver0 = 0
gen four_leaver1 = 0
gen four_leaver2 = 0
gen four_leaver3 = 0
gen four_leaver4 = 0

replace four_leaver0 = 1 if (move_time * frm_four   == 0 & move_time != .)
replace four_leaver1 = 1 if (move_time * frm_four[_n-1] == 1 & move_time != . & moved == 0)
replace four_leaver2 = 1 if (move_time * frm_four[_n-2] == 2 & move_time != . & moved == 0)
replace four_leaver3 = 1 if (move_time * frm_four[_n-3] == 3 & move_time != . & moved == 0)
replace four_leaver4 = 1 if (move_time * frm_four[_n-4] == 4 & move_time != . & moved == 0)

reg ln_wage age age2 educ2 gender four_leaver0 four_leaver1 four_leaver2 four_leaver3 four_leaver4 i.region, robust
estimates store m_left4, title(m_left4)

esttab m_left1 m_left2 m_left3 m_left4 using m_left.tex, cells(b(star fmt(3)) se(par fmt(2))) legend label varlabels(_cons constant)

* TABLE 6: by region of arrival *
```

Figure 8: Do File Also Included in the Upload