Location, location, location

David Card
Jesse Rothstein
Moises Yi

August 2022

All results have been reviewed for disclosure avoidance by the Census Bureau (DRB clearance numbers CBDRB-FY21-122, FY21-149, FY21-288, FY22-021, FY22-065, FY22-135).
Any opinions and conclusions expressed herein are those of the author(s) and do not reflect the views of the U.S. Census Bureau.

There are persistent, large differences in labor market outcomes across locations

- *Across-CZ standard deviation of mean log wages is 0.141.*
- *Controlling for observables (education, age, ethnicity):*
  - *SD of prediction is 0.050.*
  - *SD of residual is 0.109*

What accounts for these differences?

- Is it just unobserved skills of the people?
- If there is a causal effect of place?
  - Does it reflect industry clusters?
  - Is it homogenous across workers?

Mobility allows us to distinguish skills from place effects

- Most evidence on geographic differences is cross-sectional & non-causal.
- We bring panel data to bear.
  - Longitudinal Employer-Household Dynamics (LEHD) – person-by-quarter panel created by Census from unemployment insurance wage records.
  - Contains detailed geographic information, plus quarterly employment and earnings.
- Identify causal effects of place using earnings changes of movers.
  - Abowd-Kramarz-Margolis style 2-way fixed effects.
  - Person effects (= observed/unobserved skills, portable across sectors/places)
    + CZ × Industry effects (= premiums for place/industry).
  - Allows place + industry + (place × industry) pay differentials, many different forms of agglomeration effects.

Research questions

- *Are there causal place effects?*
  - ➢ Yes, but 2/3 of V[wages] across CZ's is due to "worker quality."

- *Additional contributions:*
  - ➢ Distinguishing industry and place effects
  - ➢ New perspectives on the size gradient and returns to education
  - ➢ Housing costs, place effects, and real earnings

Prior work

- Long observational literature on location wage differences.
- Limited use of movers design to study location wage differences, focusing on urban wage premia.
  - Glaeser and Mare (2001): Mover event study based on PSID & NLSY.
  - AKM models (Combes et al. 2008, Dauth et al. 2018, de la Roca and Puga 2017).
  - Several previous studies (Glaeser and Mare 2001, de la Roca and Puga 2017) emphasize potential dynamic effects.

Longitudinal Employer-Household Dynamics data

LEHD is a panel of quarterly observations on employment and earnings, with employer IDs, constructed from unemployment insurance records

- 2010Q1-2018Q2
- in quarter: must be age 22-62, earn≥$3800, 1 employer, observe 2-digit industry & CZ
- drop person if <8 quarters of earnings
- drop "transitional" quarters (first/last quarter in employment spell)
- drop very small CZ-industry cells
  - Examine 735 CZs, 24 2-digit industries. Sometimes reduce to 688 CZs in ACS or 300 w/ all industries. Also examine 4-digit industries in top 50 CZs.

Event studies of moves:

- One move: 5 quarters in one CZ-industry cell, followed by 5 quarters in another.
- Gaps of up to 4 quarters of non-employment (plus transitions) are allowed.

Additive earnings model (AKM)

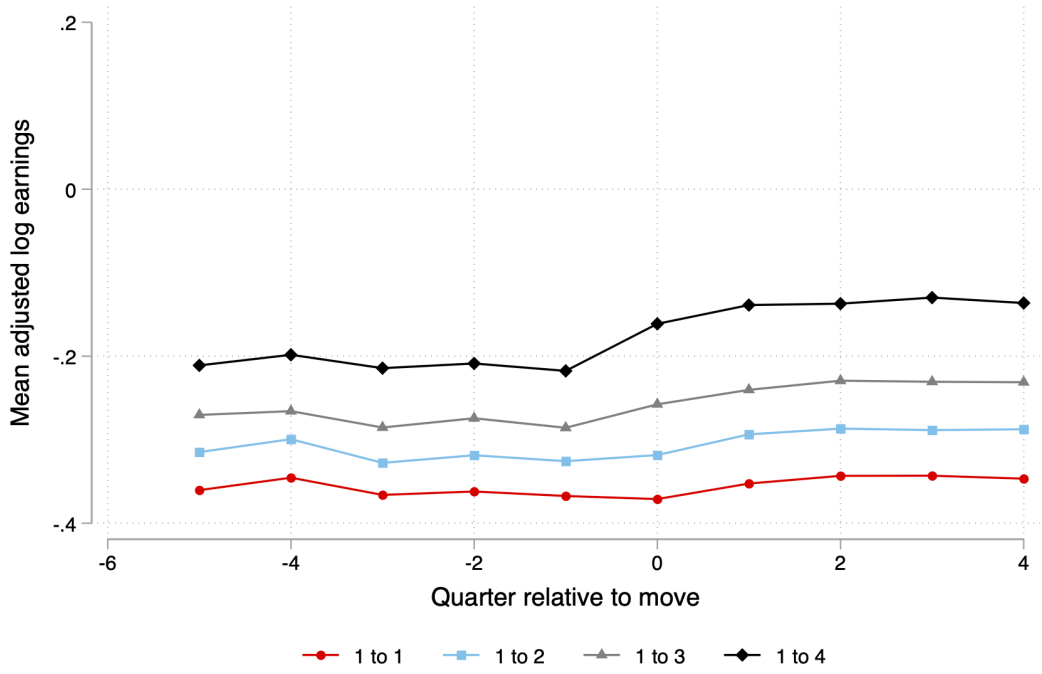
- Worker i in quarter t, working in industry-CZ cell $jc(i,t)$
- Log earnings are y_{it}.
- Decompose into permanent worker effects α_i , industry-CZ effects ψ_{jc}, and observables:

$$y_{it} = \alpha_i + \psi_{jc(i,t)} + X_{it}\beta + \varepsilon_{it}$$

- Summarize results by $V(\alpha_i)$, $V(\psi_{jc})$, $cov(\alpha_i, \psi_{jc(i,t)})$.
- Assumptions:
  - Additive separability of person and industry-CZ effects
  - Exogenous mobility – $jc(i,t)$ doesn't depend on ε_{it}, ε_{it-1}, etc.
  - Draw on tests developed by Card-Heining-Kline (2013; CHK), Card-Cardoso-Kline (2016).

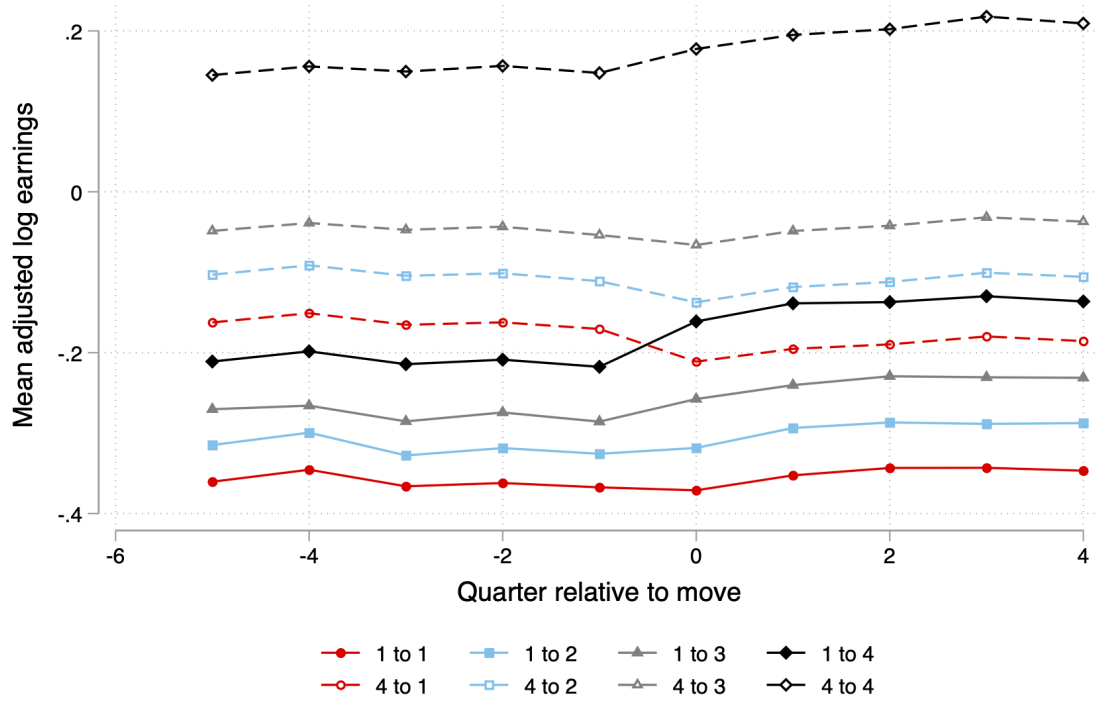
Motivation: Earnings before and after moves

- Divide CZs into quartiles (unweighted) by mean earnings.
- Event study sample
- Earnings adjusted for age, calendar time.



Motivation: Earnings before and after moves

- Divide CZs into quartiles (unweighted) by mean earnings.
- Event study sample
- Earnings adjusted for age, calendar time.



Summary of Estimated Model

Table 3: Variance Components at Individual and CZ Level

	Person-quarter level		CZ level	
	Std. Dev. or Correlation	Var. Share	Std. Dev. or Correlation	Var. Share
	(1)	(2)	(3)	(4)
Log earnings (mean at CZ level)	0.654	1.000	0.145	1.000
Variance components (std. deviations in cols 1, 3; variance shares in cols 2, 4				
Person effects	0.561	0.736	0.100	0.472
CZ-industry effects	0.097	0.022	0.064	0.196
Covariate index (Xβ)	0.150	0.053	0.005	0.001
Residual	0.243	0.138	0.000	0.000
Covariance components (correlations in cols 1, 3; variance shares in cols 2, 4				
Person/CZ-industry	0.211	0.054	0.563	0.347
Person/Covariates	-0.010	-0.004	-0.029	-0.001
CZ-industy/Covariates	0.026	0.002	-0.297	-0.009

Distinguishing the roles of CZs and industries

Big question: Are high-wage CZs just CZs with high-wage industries?

Decomposition 1: CZ effects

- The city average wage effect is $\Psi_c = \sum_j s_{jc}\psi_{jc}$.
- Let the national share of industry j be $\bar{s}_j = \sum_c \omega_c s_{jc}$, where $\omega_c = N_c/N$ is the city's population share.
- Let the *city-size weighted* (not city-industry-size weighted) industry average wage be $\bar{\psi}_j = \sum_c \omega_c \psi_{jc}$
- We can decompose Ψ_c as:

$$\Psi_c = \sum_j \bar{s}_j\bar{\psi}_j + \underbrace{\sum_j \bar{s}_j(\psi_{jc} - \bar{\psi}_j)}_{\text{City premium } (\theta_c)} + \underbrace{\sum_j (s_{jc} - \bar{s}_j)\bar{\psi}_j}_{\text{Composition: Industry specialization}} + \underbrace{\sum_j (s_{jc} - \bar{s}_j)(\psi_{jc} - \bar{\psi}_j)}_{\text{Interaction: City premium in specialty industries (=0 if no match effects)}}$$

Decomposition of across-CZ variation

Table 6: Decomposition of Variance of Average CZ Earnings Premium

		All CZ's	Top 50 CZ's Only	
		2-digit industries (1)	2-digit industries (2)	4-digit industries (3)
	Standard Dev. of Average CZ premium	0.063	0.061	0.062
	Decomposition (variance shares):			
City premium	Var(Average Earnings Premium)	1.003	0.978	0.958
Ind. specialization	Var(Composition Effect)	0.015	0.010	0.007
Interaction	Var(Interaction Effect)	0.008	0.003	0.006
	Cov(Earnings Premium, Composition Effect)	-0.027	-0.010	-0.009
	Cov(Earnings Premium, Interaction Effect)	0.002	0.025	0.039
	Cov(Composition Effect, Interaction)	0.000	-0.006	0.000

Notes: Table shows decomposition of the variance of estimated average CZ wage premium, based on equation (6) in text. Decomposition in column 1 uses main LEHD sample and 24 2-digit industries to define CZ-by-industry effects. Decompositions in columns 2 and 3 are restricted to observations in 50 largest CZ's only. Decomposition in column 2 uses 24 2-digit industries to define CZ-by-industry effects; decomposition in column 3 uses 312 4-digit industries to define CZ-by-industry effects.

Decomposing the CZ size gradient

Do big cities pay more, or attract better workers?

Decomposing the CZ size gradient

- Use the same decomposition of Ψ_c, plug back into mean wage equation:

$$\bar{y}_c = \bar{\alpha}_c + \overbrace{L_c + C_c + I_c}^{\boldsymbol{\Psi_c}} + \bar{X}_c\beta$$

  - Locational wage premium $L_c = \sum_j \bar{s}_j(\psi_{jc} - \bar{\psi}_j)$
  - Industry specialization $C_c = \sum_j (s_{jc} - \bar{s}_j)\bar{\psi}_j$
  - Interaction $I_c = \sum_j (s_{jc} - \bar{s}_j)(\psi_{jc} - \bar{\psi}_j)$

- Regress all components on ln(size) to understand relative importance

Table 8: Summary of Relationships Between CZ-level Outcomes and Log CZ Size

		LEHD	
		Estimated Coefficient (1)	Standard Error (2)
$\bar{y}_c$	***Log annual/qtrly earnings (earnings > 3800)***	0.077	(0.009)
	Basic Decomposition (All CZ's)		
$\bar{\alpha}_c$	Mean skill index / mean person effects	0.051	(0.010)
Ψ_c	CZ wage effect	0.026	(0.003)
	Percent of size effect due to skills (row 4/3)	*66.0*	
	Components of Average CZ Wage Effect (CZ's with All Industries)		
Ψ_c	CZ-average Wage Effect	0.032	(0.004)
L_c	CZ-specific premium component	0.034	(0.003)
C_c	CZ-industry composition component	-0.001	(0.000)
I_c	Interaction component	-0.001	(0.000)
	Degree of Assortative Matching within CZ (CZ's with All Industries)		
	Within-CZ skill-match correlation (correl. of person effect and industry effect)	0.061	(0.001)

Person effects are 2/3 of the size gradient

Industry does not account for any of the size gradient in CZ effects

CZ size and returns to education

Stylized fact: $\bar{y}_c^H - \bar{y}_c^L$ strongly correlated with $\ln(Size_c)$

Decompose wage gap in CZ into:

- Gap in unobserved worker skill
- Differential industry sorting
- Within-industry returns to education
- Interactions

Decomposing returns to education

- Two education groups: e=H (some college +), L (HS or less)
- Separate AKM models:
  - $y_{iet} = \alpha_i + \psi^e_{jc(i,t)} + X_{it}\beta^e + \epsilon_{iet}$
- Gap in mean earnings is:
  - $\bar{y}_{cH} - \bar{y}_{cL} = (\bar{\alpha}_{cH} - \bar{\alpha}_{cL}) + (\bar{X}_{cH}\beta^H - \bar{X}_{cL}\beta^L) + \sum_j (s_{jcH}\psi^H_{jc} - s_{jcL}\psi^L_{jc})$
  - First term is sorting on unobserved skill, second term is observables.
  - Third term is gap in CZ-by-industry effects. Similar decomposition:
    - Location wage gap (within industry): $\sum_j s_{jc}(\psi^H_{jc} - \psi^L_{jc})$
    - Differential industry sorting/shares: $\sum_j (s_{jcH} - s_{jcL})\psi_{jc}$
    - Interaction – relative clustering in industries with different premia: $\sum_j (s_{jcH} - s_{jc})(\psi^H_{jc} - \psi_{jc}) - (s_{jcL} - s_{jc})(\psi^L_{jc} - \psi_{jc})$
- Regress all components on ln(size) to understand relative importance

Table 9: Components of the Return to Education and Log CZ Size

	Estimated Coefficient (1)	Standard Error (2)
Wage gap (high- versus low-education workers)	0.0612	(0.0030)
Components of Wage Gap (column 1 = std. dev.)		
Difference in mean person effects	0.054	(.005)
Difference in covariate indexes	0.000	(.0004)
Difference in mean CZ wage effect:		
Within-industry wage gap	-0.002	(.0029)
Industry sorting	0.007	(.0007)
Interaction	0.002	(.0001)
Total	0.007	

Notes: Columns 1 and 2 show coefficient and standard error from univariate regression of CZ-specific value of wage gap term identified in row heading on the log of workforce size in CZ (estimated from ACS).

← 88% of apparent size effect on educational gap is sorting on unobservables.

Housing costs and real wages

How much of earnings premium in higher-wage / larger CZs is eaten up by housing costs?

Regressions of log housing costs on log CZ size (ACS data)

	All CZ's	Largest 50 CZ's
Housing Prices (log of home value for owners)		
Unadjusted	0.25	0.38
	(0.01)	(0.08)
Quality Adjusted	0.22	0.42
	(0.01)	(0.08)
Monthly Rent (log of rent for renters)		
Unadjusted	0.17	0.19
	(0.01)	(0.04)
Quality Adjusted	0.18	0.23
	(0.01)	(0.04)

Compare: Elasticity of earnings of college grads is 0.102, but elasticity of causal effect just 0.029.

Simple summary: move to 100 log point bigger city (1.1 M to 3 M people)

	All workers	College Educ.
Mean effect on earnings	+7.65%	+10.20%
Due to selection of workers	+ 5.05%	+7.40%
Due to causal effect of CZ	+2.60%	+2.80%
Mean effect on housing costs	+22% (housing prices, adjusted)	
	+18% (rents, adjusted)	
Mean effect on real wage (assuming 1/3 housing share)	+2.60 – 18/3 = -3.4%	+2.80 – 18/3 = -3.2%

Conclusions

- Approximately “exogenous” mobility given person effects.
- Industry & CZ effects are nearly separable; little role for industry agglomeration.
- 2/3 of across-CZ earnings differences due to sorting; 1/3 to CZ place effects.
  - Many potential sources of place effects.
  - We rule out industry specialization.
- Same for CZ size gradient: 2/3 due to skill differences.
- Higher apparent return to education in larger CZs is almost entirely due to assortative matching of highly skilled college workers to those CZs.
- Moving to a larger / higher wage CZs has a *negative* effect on earnings net of housing costs, for both college and non-college workers.