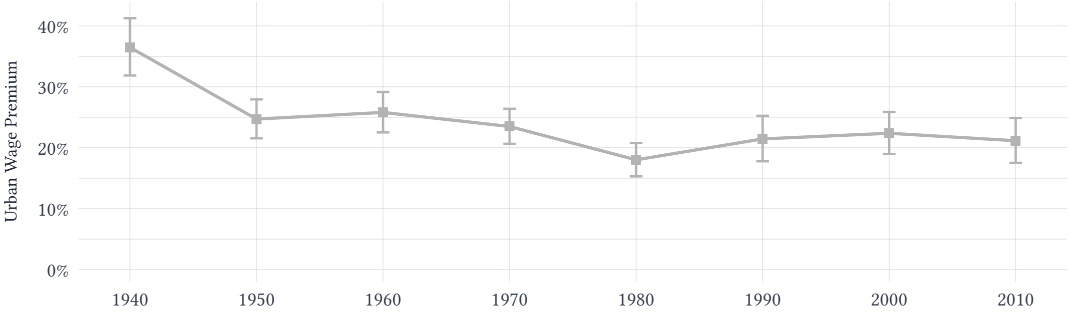
The Urban Wage Premium in Historical Perspective

Kyle Butts, Taylor Jaworski, and Carl Kitchens

December 2024



In the U.S. today, workers that live in an urban area earn ≈ 22% more than non-urban workers

- This pattern holds in cities across the world and over (at least) the last 80 years in the US (Boustan, Bunten, and Hearey, 2013)

Self-selection and the premium

While the raw data shows a large difference in earnings, it is not clear if this is a causal effect or due to non-random sorting

- College-educated and high-skilled workers are increasingly migrating to cities (Autor, 2019; Combes, Duranton, and Gobillon, 2008; Diamond, 2016; Glaeser, Kolko, and Saiz, 2001)

Indeed, panel data estimates shows the current causal wage premium is much smaller

- Glaeser and Maré (2001) estimate effects between 5 to 10%
- For blue-collar workers, Gould (2007) estimate a wage premium of only 1.2 percent

Why study the historical urban wage premium?

Over time the economic structure of a city has changed

- In the 1940s, cities were primarily manufacturing-based
- Moving towards service-sector in the 21st century (Diamond, 2016)

Understanding the wage premium over the 20th century provides insights into how the benefits of urbanity changes over a country's development

Why study the historical urban wage premium?

Over time the economic structure of a city has changed

- In the 1940s, cities were primarily manufacturing-based
- Moving towards service-sector in the 21st century (Diamond, 2016)

Understanding the wage premium over the 20th century provides insights into how the benefits of urbanity changes over a country's development

- Understanding the productivity benefits of cities in manufacture-based economies has strong implications for growth policies in developing countries

Past approach to estimating causal wage premium

The current literature focuses on a 'movers' approach to estimating the causal wage premium

Comparing individuals who enter or exit an urban location allows to control for 'worker fundamentals' via a fixed-effect strategy

- e.g. Glaeser and Maré (2001), Yankow (2006), Baum-Snow and Pavan (2012), and Card, Rothstein, and Yi (2024)

Studying the historical urban wage premium is hard

Without panel data, the standard approach is infeasible for estimating a historical causal wage premium

We use the approach of Altonji and Mansfield (2018) to estimate causal urban-wage premium in cross-sectional census data

- Leverage properties of the canonical location-choice model to control for differences in average unobservables between urban and non-urban locations

Data and Empirical Strategy

Causal Wage Premium

Heterogeneity Results

Data Sources

We use various U.S. Census surveys for each decade from 1940 to 2010:

- 1940: complete count census
- 1960–1970: 5% census samples
- 1950, 1980, 1990, and 2000: 1% census sample
- 2010: American Community Survey

We restrict the sample to

- adult males between ages 25 and 65,
- and report being employed for at least 26 weeks in the year

Observable covariates

The main outcome variable is $\log$ weekly wage

- total annual earnings / the number of weeks worked

The census data provides limited consistently measured characteristics

- age, race, educational attainment, marital status, and veteran status

Observable covariates

The main outcome variable is $\log$ weekly wage

- total annual earnings / the number of weeks worked

The census data provides limited consistently measured characteristics

- age, race, educational attainment, marital status, and veteran status

Later, we consider a measure of 'real wage' premium that adjusts for housing costs. Following Ganong and Shoag (2017), we form an individual measure of housing costs

- 5% of reported housing value or $12\times$ reported monthly rent. Divide by 48 to get the weekly housing costs

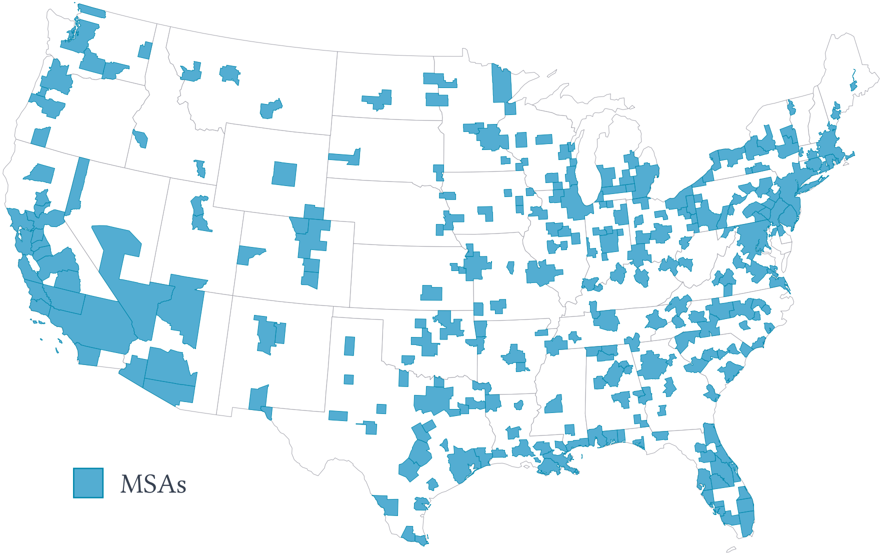
Definition of Urban Area

Our definition of ‘urban’ is defined as the Metropolitan Statistical Areas

- MSAs cover urban core and suburbs where residents live

For non-urban areas, we group together all counties not in an MSA by state

- 48 non-urban areas for people to live-in


MSAs

Empirical Strategy

We consider the following regression specifications:

$$\log(\text{wage}_i) = \beta_0 + \beta_1 \texttt{Urban}_i + \boldsymbol{W}_i \delta + u_i$$

Regressions are run separately by year and $\hat{\beta}_1$ is our wage-premium estimate for that year

- $\boldsymbol{W}_i$ are a set of worker characteristics that vary across specification

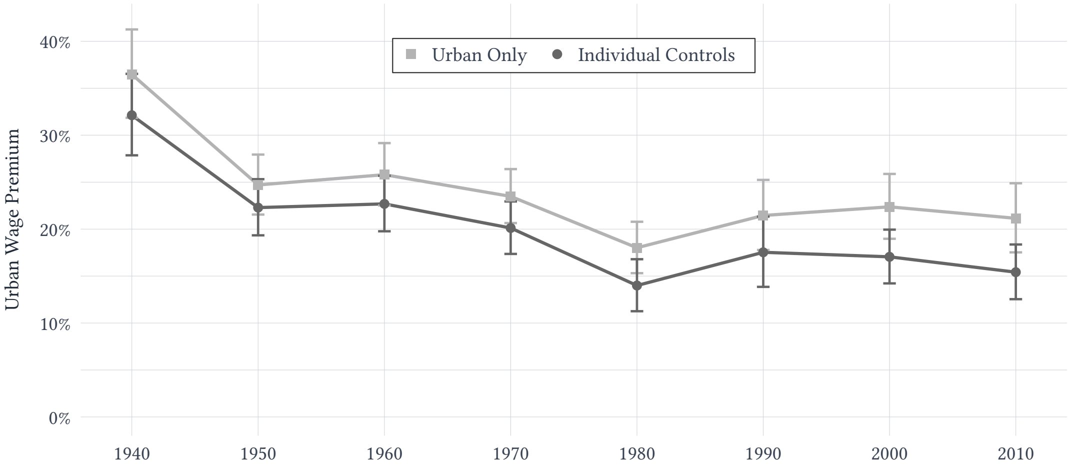
Individual level Controls

Without covariates, the raw premium (no covariates) is what we presented on the first slide

As a first pass, we control for a set of individual characteristics

- 5-year age bins
- Years of schooling
- indicator for being White

Controlling for Individual-level Characteristics



Individual level Controls

Takeaway: The individual controls absorb a small portion of the raw premium

- Much less explained than suggested by the standard “urban-living college worker” story

The estimates in later periods are much larger compared to panel data estimates

Data and Empirical Strategy

Causal Wage Premium

Heterogeneity Results

Controlling for Unobservables

In this section, we lay out the methodology to estimate the causal urban wage premium in our cross-sectional setting

- Follow closely to Altonji and Mansfield (2018)

Say individuals sort based on unobserved characteristics $\boldsymbol{X}_i^U$. The difference in $\log$ wages is conflated by differences in average $\boldsymbol{X}^U$ in urban and non-urban locations

$\implies$ It would be sufficient to control for location-specific $\bar{\boldsymbol{Z}}_{\ell(i)}$, though these are unobservable

Location-choice model

The canonical location-choice model has worker i choosing over locations, ℓ, to maximize their indirect utility

$$V_{i\ell} = \boldsymbol{W}_i \boldsymbol{A}_\ell - P_\ell + \varepsilon_{i\ell}$$

- $\boldsymbol{A}_\ell$ is a vector of location-specific amenities
- P_ℓ is the housing costs in location ℓ
- $\varepsilon_{i\ell}$ are idiosyncratic preference draws (assumed Type-1 EV)

To allow for rich sorting patterns, the willingness to pay for amenities varies by worker, $\boldsymbol{W}_i$

- E.g. college workers have preferences for certain consumptive amenities (Diamond, 2016)

Sorting by Characteristics

Since locations vary by their amenities $\boldsymbol{A}_\ell$, the kinds of individuals they attract will vary too

- In equilibrium, the location-average of observable characteristics are a function of the location's amenities $\bar{\boldsymbol{X}}_\ell = f(\boldsymbol{A}_\ell)$

Sorting by Characteristics

Since locations vary by their amenities $\boldsymbol{A}_\ell$, the kinds of individuals they attract will vary too

- In equilibrium, the location-average of observable characteristics are a function of the location's amenities $\bar{\boldsymbol{X}}_\ell = f(\boldsymbol{A}_\ell)$

While we can not observe them, the location-average of *unobservable characteristics* are also functions of amenities $\bar{\boldsymbol{X}}_\ell^U = f^U(\boldsymbol{A}_\ell)$

- The mapping and even which amenities impact sorting can differ for observable and unobservable characteristics.

Inverting the relationship

$$\bar{\boldsymbol{X}}_\ell = f(\boldsymbol{A}_\ell) \text{ and } \bar{\boldsymbol{X}}_\ell^U = f^U(\boldsymbol{A}_\ell)$$

Say we knew the mappings$f$ and f^U. Under conditions we will list below, we can 'invert' observable averages to learn about amenities

$$\boldsymbol{A}_\ell = f^{-1}(\bar{\boldsymbol{X}}_\ell)$$

Inverting the relationship

$$\bar{\boldsymbol{X}}_\ell = f(\boldsymbol{A}_\ell) \text{ and } \bar{\boldsymbol{X}}_\ell^U = f^U(\boldsymbol{A}_\ell)$$

Say we knew the mappings f and f^U. Under conditions we will list below, we can 'invert' observable averages to learn about amenities

$$\boldsymbol{A}_\ell = f^{-1}(\bar{\boldsymbol{X}}_\ell)$$

Then apply f^U to learn about unobservables

$$\bar{\boldsymbol{X}}_\ell^U = f^U(f^{-1}(\bar{\boldsymbol{X}}_\ell))$$

Inverting the relationship

$$\bar{\boldsymbol{X}}_\ell^U = f^U(f^{-1}(\bar{\boldsymbol{X}}_\ell))$$

Under this set of assumptions, the location choice model implies that $f^U(f^{-1}(\cdot))$ is actually a *linear mapping*, $\bar{\boldsymbol{X}}_\ell^U = \bar{\boldsymbol{X}}_\ell \Pi$

Inverting the relationship

$$\bar{\boldsymbol{X}}_\ell^U = f^U(f^{-1}(\bar{\boldsymbol{X}}_\ell))$$

Under this set of assumptions, the location choice model implies that $f^U(f^{-1}(\cdot))$ is actually a *linear mapping*, $\bar{\boldsymbol{X}}_\ell^U = \bar{\boldsymbol{X}}_\ell \Pi$

$\implies$ Upshot: Controlling *linearly* for $\bar{\boldsymbol{X}}_\ell$ in our wage premium model is sufficient to control for differences in unobservables between urban and non-urban locations

Necessary Assumptions

The first three assumptions are standard:

A1 The indirect utility function is $V_{i\ell} = \boldsymbol{W}_i \boldsymbol{A}_\ell - P_\ell + \varepsilon_{i\ell}$

A2 $\varepsilon_{i\ell}$ are iid type-1 Extreme Value

$\rightarrow$ standard in urban/trade literature

A3 When choosing locations, workers take prices and amenities as given

$\rightarrow$ They do not internalize their own effect on prices and amenities

Necessary Assumptions

The next two assumptions restrict the relationship between observables, unobservables, and amenities. Remember our mapping relies on:

Observables teach us about amenities

A4 The function $A_\ell = f(\bar{X}_\ell)$ is invertible

$\rightarrow$ For each amenity, we need some X_i that explains differences in WTP for that amenity

Necessary Assumptions

The next two assumptions restrict the relationship between observables, unobservables, and amenities. Remember our mapping relies on:

Observables teach us about amenities

A4 The function $A_\ell = f(\bar{X}_\ell)$ is invertible

- $\rightarrow$ For each amenity, we need some X_i that explains differences in WTP for that amenity

Amenities teach us about unobservables

A5 The set of location-specific amenities that drive sorting on observables contains all amenities that drive sorting based on unobservable covariates

Result from Altonji and Mansfield (2018)

With assumptions A1-A5, the location choice model implies the mapping between $\bar{X}_\ell$ and $\bar{X}_\ell^U$ is linear

We leverage this to isolate the *causal* urban wage premium by modifying our estimation equation to include a set of location-averages of observable characteristics

$$\log(\text{wage}_i) = \beta_0 + \beta_1 \texttt{Urban}_i + \boldsymbol{W}_i \delta + \bar{\boldsymbol{X}}_\ell \gamma + u_i$$

Group Averages

For our group-averages, we include variables we think explain differences in WTP for amenities:

- the share of white workers (the share of non-white workers drops out)
- share of residents in each 5-year age bin
- share of veterans
- share of married and together, married and apart, and single
- share with $<$ HS, HS degree, some college, BA degree, and post-secondary education
- a measure of market access calculated as in Jaworski and Kitchens (2019)

Plausibility of A4

While there are theoretically many amenities people can ‘shop with their feet for’, many are tightly bundled together

- Using a principle components analysis on a wide array of amenities, Diamond (2016) finds 5 ‘bundles’ of amenities across locations

Therefore, we need the effective rank of our group averages to be 5 or larger

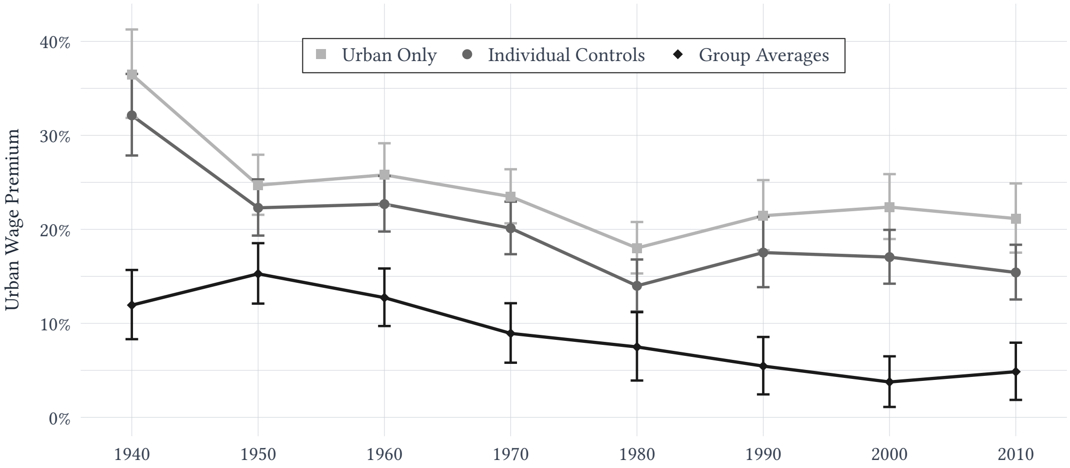
- Running a similar PCA, we find 6 components explain more than 5 percent of the variance across years

Plausibility of A5

A5 is a more difficult assumption to assess since it is an assumption about unobservable characteristics

- Later, we think our ability to match estimates in 2000 and 2010 help alleviate concerns about failing to control for unobservables

Causal Urban Wage Premium Estimates



Takeaways

From 1990–2010, the causal premium estimate is around 5%, in line with results from Glaeser and Maré (2001) and Gould (2007)

- This gives us confidence in our ability to control for sorting based on unobservables

Takeaways

From 1990–2010, the causal premium estimate is around 5%, in line with results from Glaeser and Maré (2001) and Gould (2007)

- This gives us confidence in our ability to control for sorting based on unobservables

Looking into past decades, the causal estimates grow over time, peaking between 12 and 15% in the middle of the 20th century

In line with the historical evidence:

- In middle 20th century, manufacturing bases in cities offering higher wages
- Decline in wage premium concurrent with sub-urbanization

Wages net of Housing Costs

Of course, urban areas also face higher housing costs

- Higher wages workers receive might be extracted by landlords via higher rent

Wages net of Housing Costs

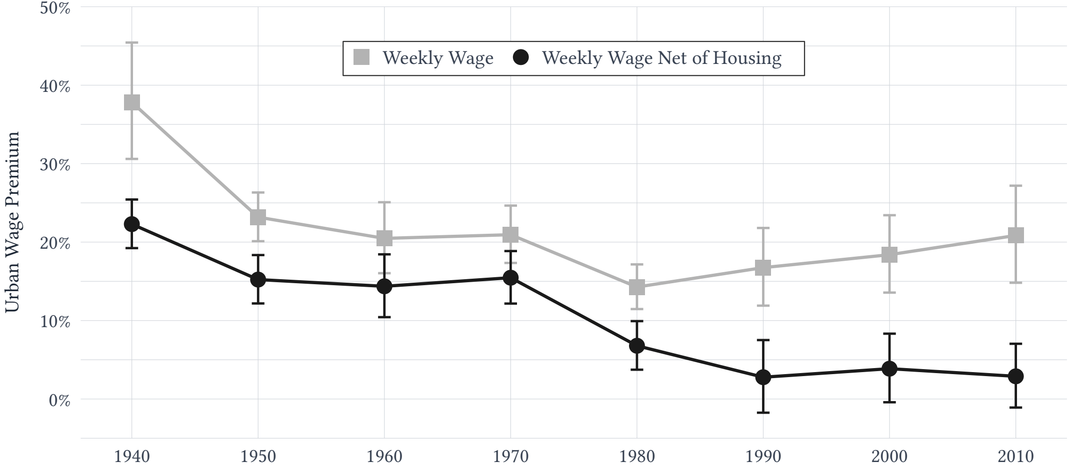
Of course, urban areas also face higher housing costs

- Higher wages workers receive might be extracted by landlords via higher rent

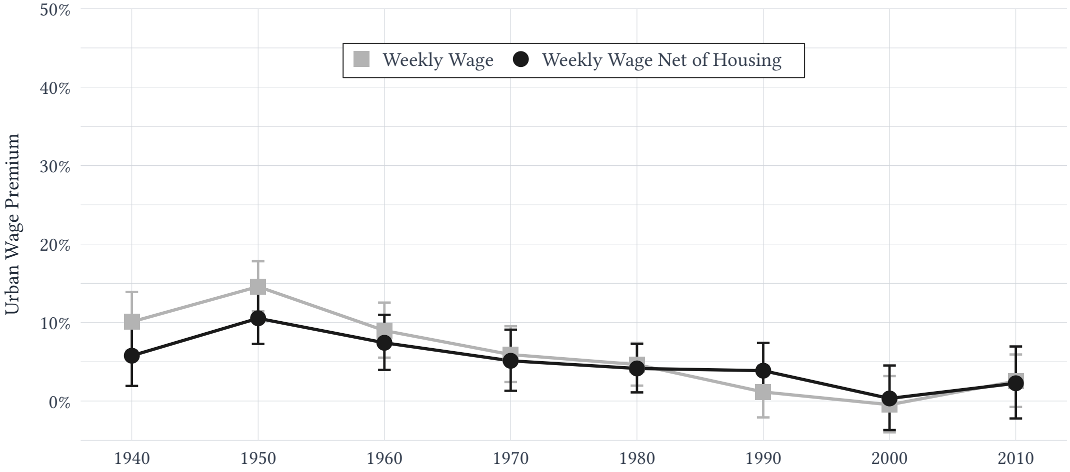
We rerun our analysis using $\log(\text{wage}_i - P_\ell)$

- Following Ganong and Shoag (2017), 5% of reported housing value or 12× reported monthly rent. Divide by 48 to get the weekly housing costs
  - → Averaged over individuals to abstract from individual preferences for housing
  - → The 1950 census had no housing estimates, so we use an aggregate housing cost measure from the Census

Raw Urban (Net of Housing) Wage Premium Estimates



Causal Urban (Net of Housing) Wage Premium Estimates



Data and Empirical Strategy

Causal Wage Premium

Heterogeneity Results

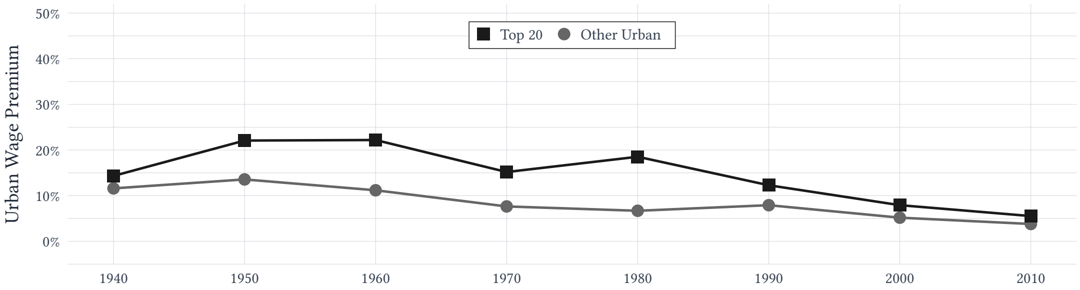
Heterogeneity by City Size

Next, we explored heterogeneity in the causal wage premium.

First, the literature has shown that larger cities have a larger causal premium (Baum-Snow and Pavan, 2012)

- We interact the urban indicator with being a top-20 city in terms of population (based on 1940 population)

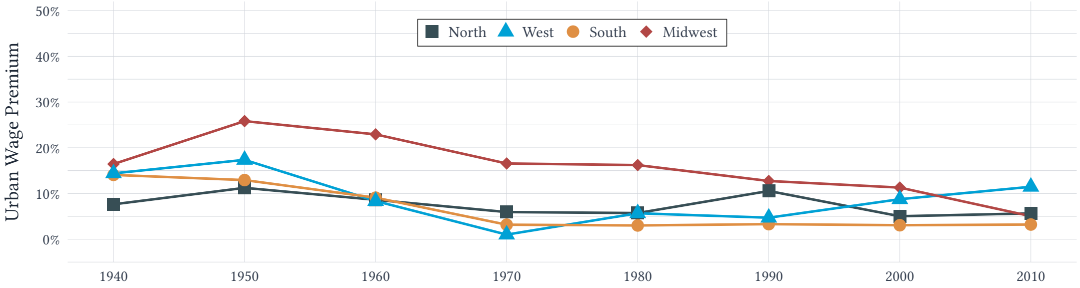
Heterogeneity by City Size



Indeed we see the wage premium is moderately larger across years (Baum-Snow and Pavan, 2012), but the difference has shrunk in recent decades

- Consistent with the theory of sorting being increasingly important

Heterogeneity by Census Region



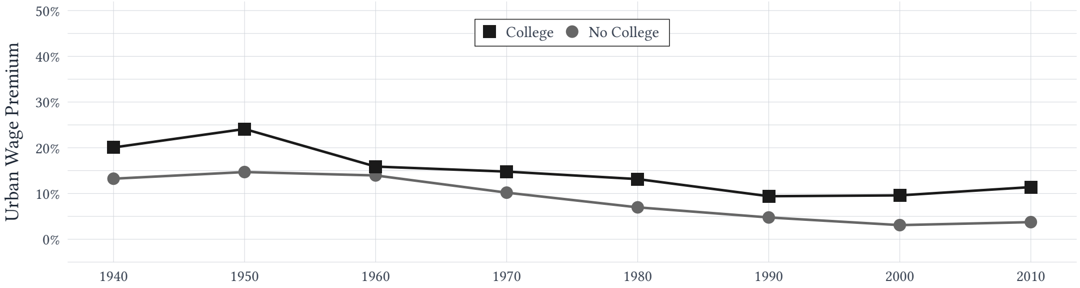
- The Midwest had a much higher wage premium, consistent with cities' historical premium being driven by manufacturing
- A recent rise in the West's premium, consistent with technology sector

Heterogeneity for groups of workers

Next, we present three forms of worker-heterogeneity: college education and race

- For each, we take include separate group-averages by race/education because the sorting patterns (and hence linear mappings) might differ by these categories

Heterogeneity by College Education

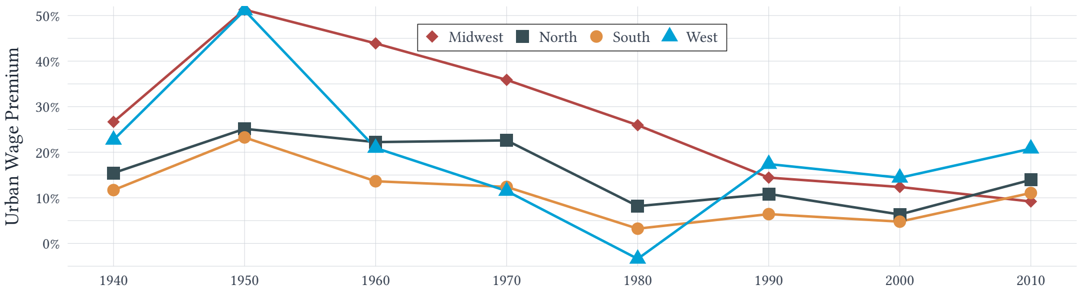


The wage-premium is smaller for non-college educated workers, consistent with (Gould, 2007)

Heterogeneity by Race

The relatively higher wage premium by Black workers occurs during the period of the Great Migration (Boustan, 2016)

Regional Heterogeneity for Black Workers



At risk of stretching the data too thin, can see the Midwest and Western premiums served as a 'pull factor' for the Great Migration

Summary

Estimate a causal urban wage premium from 2010 back until 1940

- The causal premium was between 12 and 15% in 1940–1960 and has shrunk over time to around 5%

Heterogeneity analysis in line with historical narrative:

- High-manufacturing in the Midwest gave high wage premia, particularly for Black workers
- College-educated workers benefit more from urbanity

These result help us understand how agglomeration benefits accrue to workers at different points in economic development

References I

Altonji, Joseph G. and Richard K. Mansfield (2018). "Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects". *American Economic Review* 108.10, pp. 2902–46.

Autor, David H. (May 2019). "Work of the Past, Work of the Future". *AEA Papers and Proceedings* 109, pp. 1–32. DOI: 10.1257/pandp.20191110.

Baum-Snow, Nathaniel and Ronni Pavan (2012). "Understanding the City Size Wage Gap". *Review of Economic Studies* 79.1, pp. 88–127.

Boustan, Leah Platt (2016). *Competition in the Promised Land*. Princeton University Press.

Boustan, Leah Platt, Devin Michelle Bunten, and Owen Hearey (May 2013). *Urbanization in the United States, 1800-2000*. Working Paper 19041. National Bureau of Economic Research. DOI: 10.3386/w19041.

Card, David, Jesse Rothstein, and Moises Yi (2024). *Location, Location, Location*. Tech. rep. Working Paper.

References II

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2008). "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics* 63.2, pp. 723–742.

Diamond, Rebecca (Mar. 2016). "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000". *American Economic Review* 106.3, pp. 479–524. ISSN: 0002-8282. DOI: 10.1257/aer.20131706.

Ganong, Peter and Daniel Shoag (2017). "Why Has Regional Income Convergence in the US Eeclined?" *Journal of Urban Economics* 102, pp. 76–90.

Glaeser, Edward L., Jed Kolko, and Albert Saiz (2001). "Consumer City". *Journal of Economic Geography* 1.1, pp. 27–50.

Glaeser, Edward L. and David C. Maré (Apr. 2001). "Cities and Skills". *Journal of Labor Economics* 19.2, pp. 316–342. ISSN: 0734-306X, 1537-5307. DOI: 10.1086/319563.

References III

Gould, Eric D. (2007). "Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium". *Review of Economic Studies* 74.2, pp. 477–506.

Jaworski, Taylor and Carl T. Kitchens (2019). "National Policy for Regional Development: Historical Evidence from Appalachian Highways". *Review of Economics and Statistics* 101.5, pp. 777–790. DOI: 10.1162/rest_a_00808.

Yankow, Jeffrey (2006). "Why Do Cities Pay More? An Empirical Examination of Some Competing Theories of the Urban Wage Premium". *Journal of Urban Economics* 60.2, pp. 139–161.