Spatial Wage Estimates for 19th Century US Workers

Jason Dunn

Department of Economics, Boston University

November 21, 2026

Introduction

Motivation/Literature

- Popular questions in economic history: **inequality, economic mobility**
- Problem: Difficult to measure socioeconomic status (**SES**) in history
  - Limited info on individual income, wealth, etc.
- Standard proxy: **occupational scores**. But limited in variation/accuracy. Improvements:
  - (Song et al., 2020); (Ward, 2023): Literacy shares within occupation-race-region
  - (Saavedra and Twinam, 2020): LLM on informative correlates
  - (Collins and Zimran, 2023): wealth data, 1850-1870 censuses

Introduction

Motivation/Literature

- Popular questions in economic history: **inequality, economic mobility**
- Problem: Difficult to measure socioeconomic status (**SES**) in history
  - Limited info on individual income, wealth, etc.
- Standard proxy: **occupational scores**. But limited in variation/accuracy. Improvements:
  - (Song et al., 2020); (Ward, 2023): Literacy shares within occupation-race-region
  - (Saavedra and Twinam, 2020): LLM on informative correlates
  - (Collins and Zimran, 2023): wealth data, 1850-1870 censuses

This paper

- **Core contribution:** tightly identified "spatial wages" for **manufacturing** workers
  - COM Firm-level data (Hornbeck et al., 2024)
  - Supplemented with GPIH occupational quotes for non-manufacturing workers
- Combine with literacy & wealth $\rightarrow$ composite SES; validate on 1880
  - Implications for mobility (Ward, 2023) & inequality (Atack et al., 2004)
- Creating a pipeline for validating/benchmarking SES measures

Outline for Today

- Data
- Methodology
  - Manufacturing spatial wages (panel, 1850–1880)
  - Coarser wage supplement for broader coverage
  - Validation on aggregate cells
  - Construction of Composite SES Index
- Results
  - Spatial wage tracks aggregate comparables
  - Composite matches/improves income proxy vs. existing SES scores
  - Mobility lower; inequality greater increase
- Discussion/Next Steps

Data

Data

Wage estimates

1. 1850–1880 Census of Manufactures (**COM**) (Hornbeck et al., 2024) sample
   - Establishments $>$ \$500: township, industry, wages, workers wages summ map
2. 1850–1880 Census of Population (**COP**) summ
3. 1850–1880 Census Place Project (**CPP**) (Berkes et al., 2023) map
4. Global Price and Income History (**GPIH**) quotes
   - Farm/common labor, domestics, clerks, crafts, miners, teachers, ...
   - Typically state / region / US—not local firm geography

Data

Wage estimates

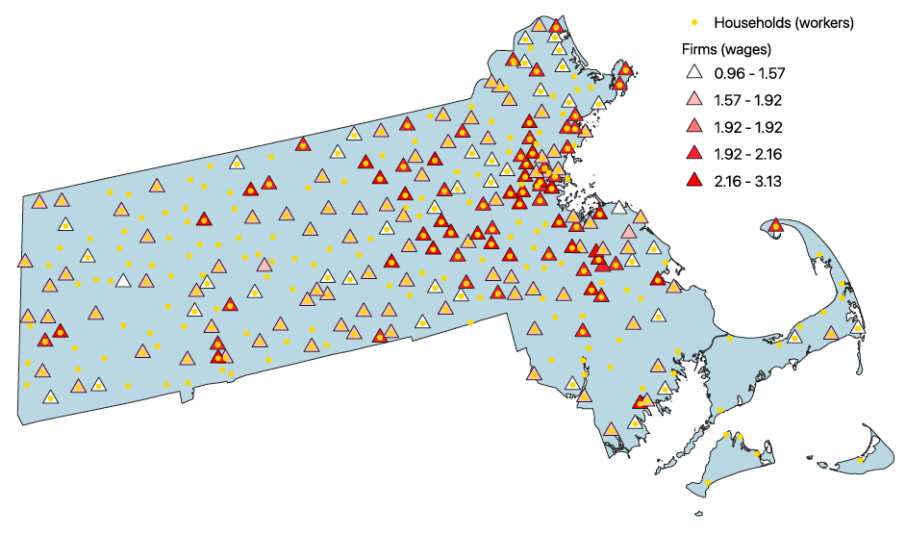
1. 1850–1880 Census of Manufactures (**COM**) (Hornbeck et al., 2024) sample
   - Establishments $>$ \$500: township, industry, wages, workers wages summ map
2. 1850–1880 Census of Population (**COP**) summ
3. 1850–1880 Census Place Project (**CPP**) (Berkes et al., 2023) map
4. Global Price and Income History (**GPIH**) quotes
   - Farm/common labor, domestics, clerks, crafts, miners, teachers, ...
   - Typically state / region / US—not local firm geography

Validation/Benchmarking

5. 1888-1890 Cost of Living Survey (**COLS**): 6,627 HH incomes incomes summ
6. Department of Labor (**DOL**): 146 city-occupation wages, 1870–1898 map
7. Historical Labor Statistics Project (**HLSP**): $\sim$70k worker wages/incomes summ
8. 1915 Iowa Census ($\sim$20k incomes) and 1940 Full Count Census ($\sim$38m incomes)
9. Other SES scores: Song, Adj. Song, Wealth, LIDO, Occscore, ...

Methodology

Methodology: Thought Experiment



Methodology: Spatial Wages

Core method — panel for each COM decade 1850–1880 For **manufacturing** worker i at location l, industry j, demo g, skill s:

$$\omega_{i,l,j,g,s} = \frac{\sum_{f \in F_{l,j,d}} (W_{f,j,g,s} \cdot L_{f,j,g})}{\sum_{f \in F_{l,j,d}} L_{f,j,g}} \tag{1}$$

- $d = 15$ miles; employment weights $L_{f,j,g}$
- Fallbacks: demo/skill fill $\rightarrow$ county $\rightarrow$ state ($\frac{1}{4}$ precision weight)

Methodology: Spatial Wages

Core method — panel for each COM decade 1850–1880 For **manufacturing** worker i at location l, industry j, demo g, skill s:

$$\omega_{i,l,j,g,s} = \frac{\sum_{f \in F_{l,j,d}} (W_{f,j,g,s} \cdot L_{f,j,g})}{\sum_{f \in F_{l,j,d}} L_{f,j,g}} \tag{1}$$

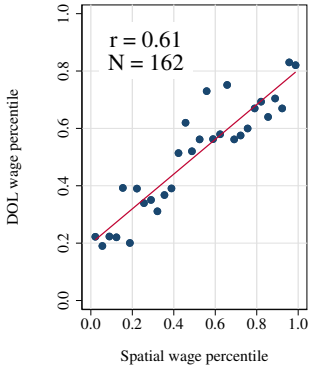
- $d = 15$ miles; employment weights $L_{f,j,g}$
- Fallbacks: demo/skill fill $\rightarrow$ county $\rightarrow$ state ($\frac{1}{4}$ precision weight)

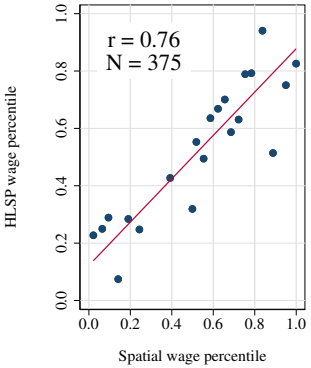
Supplemental wages — GPIH occupational quotes at state/region/US level for non-manufacturing workers coverage

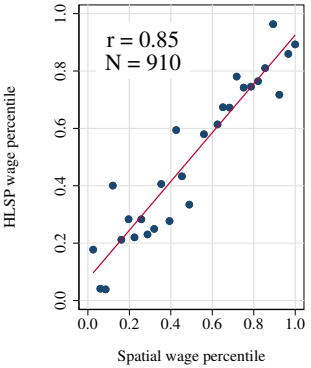
- Tradeoff: broader coverage $\neq$ local identification

Results

Wage Validation







Wage Comparison

	DOL, 1870–1880	HLSP manufacturing	HLSP all workers
Spatial wage	0.61	0.76	0.85
Occscore	0.27	0.69	0.80
Occupational earnings	0.27	0.69	0.81
LIDO	0.20	0.76	0.84
Nam-Powers-Boyd	0.06	0.06	0.64
Adjusted Song	0.04	0.54	0.39
Collins-Zimran wealth	0.01	-0.66	-0.25
Siegel prestige	-0.10	0.76	0.36
Occupational education	-0.12	0.76	0.45
Duncan SEI	-0.15	0.43	0.59
Song	-0.20	0.61	0.44
Cells	162	375	910
Workers	–	12,236	24,483

Methodology: Constructing Composite SES

- Three components
  1. Spatial wage
  2. Individual Literacy (Song et al., 2020); (Ward, 2023)
  3. Local average wealth (Collins and Zimran, 2023)
- Local CPP level, demographic (white male, non-white male, female), occupation
  - Bayes' shrinkage on wealth to region average
- Optimal weighting
  - Maximize correlation with **COLS** data
  - Optimal weights: 0.51 wage / 0.40 literacy / 0.09 wealth
- Out of sample fit: **HLSP** (and 1915 Iowa, 1940 Full Count Census)

$$\widetilde{y}_i = \alpha + \beta \widetilde{S}_i + \varepsilon_i. \tag{2}$$

SES Variance Decomposition

	Occupation	+ Region	+ State	+ Place	+ Demo/Skill
Composite SES	0.44	0.25	0.05	0.12	0.14
Equal weights	0.40	0.24	0.04	0.19	0.13
Spatial wage	0.85	0.02	0.04	0.05	0.04
Literacy share	0.18	0.37	0.06	0.17	0.22
Average wealth	0.42	0.09	0.03	0.44	0.02
Adjusted Song	0.58	0.36	0.00	0.01	0.05
LIDO	0.66	0.15	0.07	0.02	0.10
Collins-Zimran wealth	0.73	0.13	0.14	0.00	0.00
Song	0.98	0.00	0.00	0.00	0.02
Siegel prestige	1.00	0.00	0.00	0.00	0.00
Nam-Powers-Boyd	1.00	0.00	0.00	0.00	0.00
Duncan SEI	1.00	0.00	0.00	0.00	0.00
Occscore	1.00	0.00	0.00	0.00	0.00
Occupational education	1.00	0.00	0.00	0.00	0.00
Occupational earnings	1.00	0.00	0.00	0.00	0.00

Composite SES Results

	COLS	HLSP		Iowa 1915		1940	
	training	mfg.	all	mfg.	all	mfg.	all
Composite SES	0.56	0.75	0.75	0.25	0.28	0.35	0.35
Equal weights	0.53	0.76	0.76	0.17	0.31	0.36	0.36
Spatial wage	0.38	0.78	0.85	0.68	0.40	0.30	0.30
Literacy share	0.42	0.12	0.42	-0.51	-0.38	0.18	0.18
Average wealth	0.29	0.46	0.46	0.14	0.29	0.30	0.30
Adjusted Song	0.51	0.54	0.42	-0.28	-0.08	0.33	0.33
Siegel prestige	0.49	0.72	0.37	-0.24	-0.13	0.26	0.26
Collins-Zimran wealth	0.47	-0.49	-0.22	-0.05	0.09	0.25	0.25
LIDO	0.44	0.79	0.85	0.17	0.15	0.62	0.62
Nam-Powers-Boyd	0.42	0.26	0.67	-0.09	-0.30	0.41	0.41
Duncan SEI	0.38	0.48	0.62	-0.20	-0.37	0.37	0.37
Occscore	0.37	0.73	0.81	-0.06	-0.13	0.38	0.38
Occupational education	0.31	0.76	0.48	-0.20	-0.28	0.27	0.27
Occupational earnings	0.31	0.73	0.81	-0.10	-0.15	0.31	0.31
Song	0.26	0.59	0.46	-0.26	-0.08	0.21	0.21
Cells	170	546	919	31	34	81,106	81,106
Households / workers	4,121	14,625	24,721	3,730	15,987	3,815,030	3,815,030

Mobility and Inequality

Methodology: Mobility and Inequality

- Mobility
  - Replicate and use composite index in results of (Ward, 2023)
    - Instrument for father's occupation with a second observation of that father's occupation in an adjacent census, to clean measurement error
    - Weight linked data based on representativeness
    - Include southern blacks in 1870-1880 by imputing their fathers as slaves

$$\text{Rank}_{i,c}^{\text{son}} = \alpha_c + \beta_c \widehat{\text{Rank}}_{f,c}^{\text{father}} + \varepsilon_{i,c} \tag{3}$$

$$\text{Rank}_{f,c}^{\text{father}} = \gamma_c + \delta_c \text{Rank}_{f,c,t+/-1}^{\text{father}} + \nu_{f,c} \tag{4}$$

Methodology: Mobility and Inequality

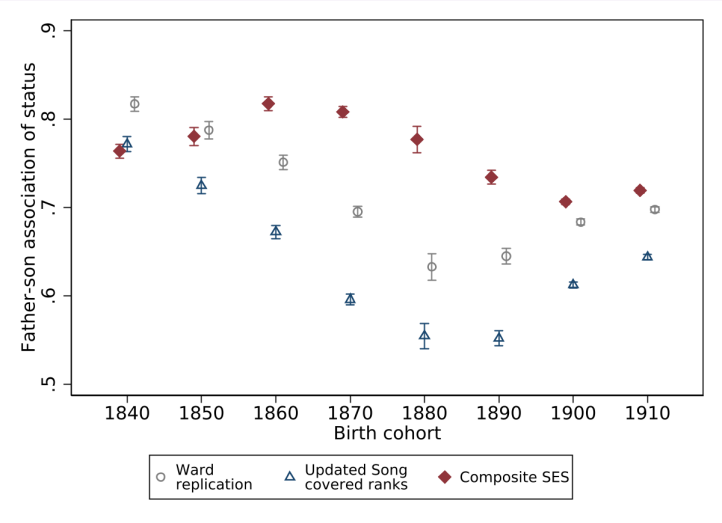
- Mobility
  - Replicate and use composite index in results of (Ward, 2023)
    - Instrument for father's occupation with a second observation of that father's occupation in an adjacent census, to clean measurement error
    - Weight linked data based on representativeness
    - Include southern blacks in 1870-1880 by imputing their fathers as slaves

$$\text{Rank}_{i,c}^{\text{son}} = \alpha_c + \beta_c \widehat{\text{Rank}}_{f,c}^{\text{father}} + \varepsilon_{i,c} \tag{3}$$

$$\text{Rank}_{f,c}^{\text{father}} = \gamma_c + \delta_c \text{Rank}_{f,c,t+/-1}^{\text{father}} + \nu_{f,c} \tag{4}$$

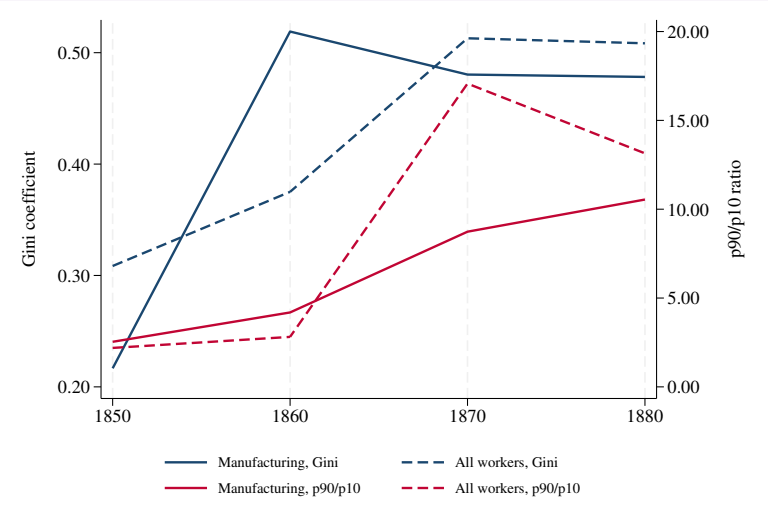
- Inequality
  - Gini and p90/p10 of composite vs. other SES scores
  - Report **manufacturing** (spatial wage) and **all workers** (COM + GPIH)

Intergenerational Mobility Results



▸ Robustness ▸ SES

Inequality Results: Composite SES



▸ Components

Inequality Results: All SES (All Worker Gini)

	1850	1860	1870	1880	1880−1850
Composite SES	0.31	0.38	0.51	0.51	0.20
Equal weights	0.30	0.57	0.52	0.50	0.20
Spatial wage	0.41	0.50	0.52	0.57	0.16
Average wealth	0.40	0.75	0.61	0.60	0.19
Occupational education	0.46	0.50	0.46	0.50	0.04
LIDO	0.55	0.61	0.61	0.60	0.06
Occupational earnings	0.54	0.59	0.59	0.60	0.06
Nam-Powers-Boyd	0.56	0.63	0.64	0.65	0.09
Occscore	0.49	0.58	0.58	0.60	0.11
Adjusted Song	0.51	0.57	0.62	0.63	0.11
Duncan SEI	0.55	0.62	0.63	0.67	0.12
Collins-Zimran wealth	0.44	0.51	0.54	0.57	0.13
Siegel prestige	0.40	0.51	0.54	0.54	0.14
Song	0.40	0.49	0.55	0.55	0.15

▸ Plot ▸ p90/p10

Conclusion

Currently . . .

- Develop a local method for **manufacturing** wages (panel 1850–1880; validate on 1880)
- Supplement with GPIH for all-worker results (coverage up, geographic precision down)
- Composite on wage + literacy + wealth matches/improves income proxy vs. existing SES scores
- Implications: higher measured inequality; lower intergenerational mobility

Conclusion

Currently . . .

- Develop a local method for **manufacturing** wages (panel 1850–1880; validate on 1880)
- Supplement with GPIH for all-worker results (coverage up, geographic precision down)
- Composite on wage + literacy + wealth matches/improves income proxy vs. existing SES scores
- Implications: higher measured inequality; lower intergenerational mobility

Hopefully coming . . .

- Refining construction and validation pipeline
- Subset of population with individual SES?
  - Link on name to census

Conclusion

Currently . . .

- Develop a local method for **manufacturing** wages (panel 1850–1880; validate on 1880)
- Supplement with GPIH for all-worker results (coverage up, geographic precision down)
- Composite on wage + literacy + wealth matches/improves income proxy vs. existing SES scores
- Implications: higher measured inequality; lower intergenerational mobility

Hopefully coming . . .

- Refining construction and validation pipeline
- Subset of population with individual SES?
  - Link on name to census

Thank You!

jtdunn@bu.edu

References I

Atack, J., F. Bateman, and R. A. Margo (2004). Skill Intensity and Rising Wage Dispersion in Nineteenth-Century American Manufacturing. *The Journal of Economic History 64*(1), 172–192.

Berkes, E., E. Karger, and P. Nencka (2023, January). The census place project: A method for geolocating unstructured place names. *Explorations in Economic History 87*, 101477.

Collins, W. J. and A. Zimran (2023, July). Working Their Way Up? US Immigrants' Changing Labor Market Assimilation in the Age of Mass Migration. *American Economic Journal: Applied Economics 15*(3), 238–269.

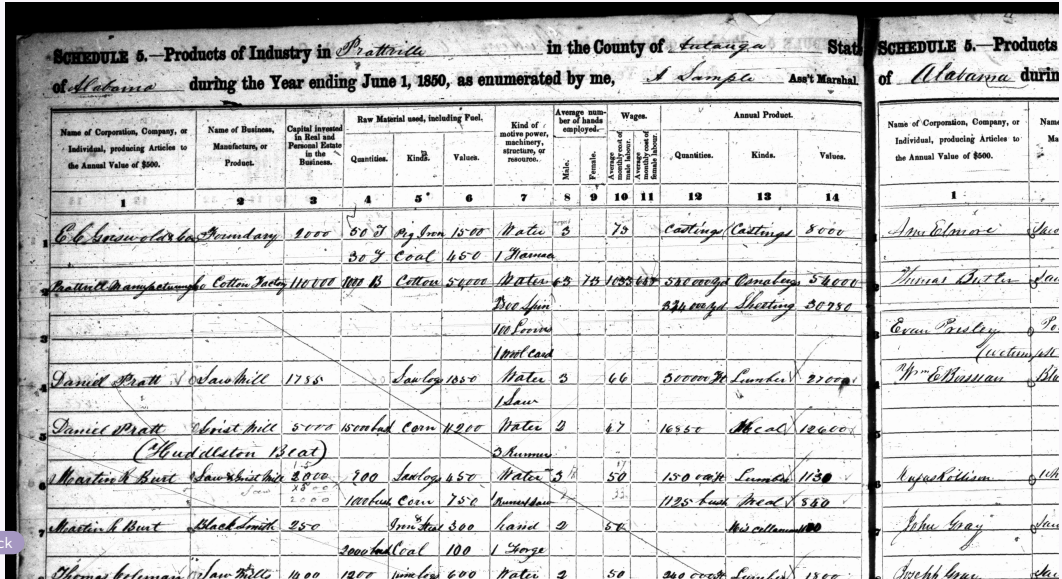
Hornbeck, R., S. Hsu, A. Humlum, and M. Rotemberg (2024, April). Gaining Steam: Incumbent Lock-in and Entrant Leapfrogging.

Saavedra, M. and T. Twinam (2020, January). A machine learning approach to improving occupational income scores. *Explorations in Economic History 75*, 101304.

Song, X., C. G. Massey, K. A. Rolf, J. P. Ferrie, J. L. Rothbaum, and Y. Xie (2020, January). Long-term decline in intergenerational mobility in the United States since the 1850s. *Proceedings of the National Academy of Sciences 117*(1), 251–258.

Ward, Z. (2023, December). Intergenerational Mobility in American History: Accounting for Race and Measurement Error. *American Economic Review 113*(12), 3213–3248.

Appendix: COM Page Sample



SCHEDULE 5.—Products of Industry in Prattville in the County of Autauga State of Alabama during the Year ending June 1, 1850, as enumerated by me, A Sample Ass't Marshal.

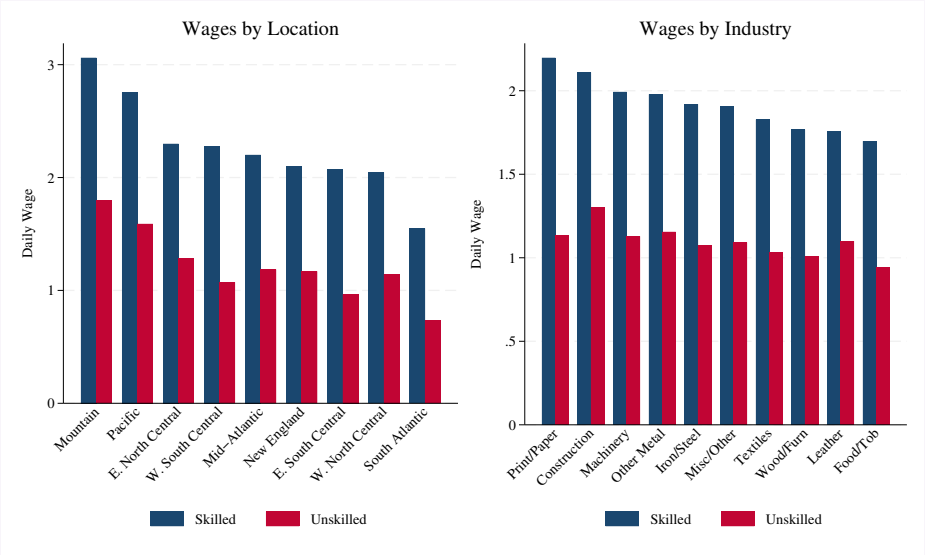
Name of Corporation, Company, or Individual, producing Articles to the Annual Value of $500.	Name of Business, Manufacture, or Product.	Capital invested in Real and Personal Estate in the Business.	Raw Material used, including Fuel. Quantities.	Kinds.	Values.	Kind of motive power, machinery, structure, or resource.	Average number of hands employed. Male.	Female.	Wages. Average monthly cost of male labour.	Average monthly cost of female labour.	Annual Product. Quantities.	Kinds.	Values.
1	2	3	4	5	6	7	8	9	10	11	12	13	14
E C Griswold & Co	Foundary	9000	50 T	Pig Iron	1500	Water	3		75		Castings	Castings	8000
			30 T	Coal	450	1 Hamer							
Prattville Manufacturing Co	Cotton Factory	110000	1000 B	Cotton	50000	Water	63	73	1033	667	540000 yd	Osnaburg	54000
						300 Spin					324000 yd	Sheeting	30780
						100 Looms							
						1 Wool Card							
Daniel Pratt	Saw Mill	1785		Saw logs	1850	Water	3		66		300000 ft	Lumber	27000
						1 Saw							
Daniel Pratt	Grist Mill	5000	15000 bush	Corn	11200	Water	2		47		16850	Meal	12600
(Huddleston Beat)						3 Runners							
Martin K Burt	Saw & Grist Mill	2000	900	Saw logs	450	Water	3		50		150000 ft	Lumber	1130
		2000	1000 bush	Corn	750	1 Runner & Saw			33		1125 bush	Meal	840
Martin K Burt	Black Smith	250		Iron & Steel	300	hand	2		50			Mis collaneous	100
			2000 bush	Coal	100	1 Forge							
Thomas Coleman	Saw Mills	1400	1200	Pine logs	600	Water	2		50		240000 ft	Lumber	1800

SCHEDULE 5.—Products of Alabama durin

Name of Corporation, Company, or Individual, producing Articles to the Annual Value of $500.	Name Ma
1	
Ann Elmore	Sau
Thomas Butler	Sau
Evan Presley	Po
Wm E Bassman	Bla
Hugh Robinson	Wh
John Gray	Sau
Joseph Gray	Sau

back

Appendix: 1880 COM Wage Distribution



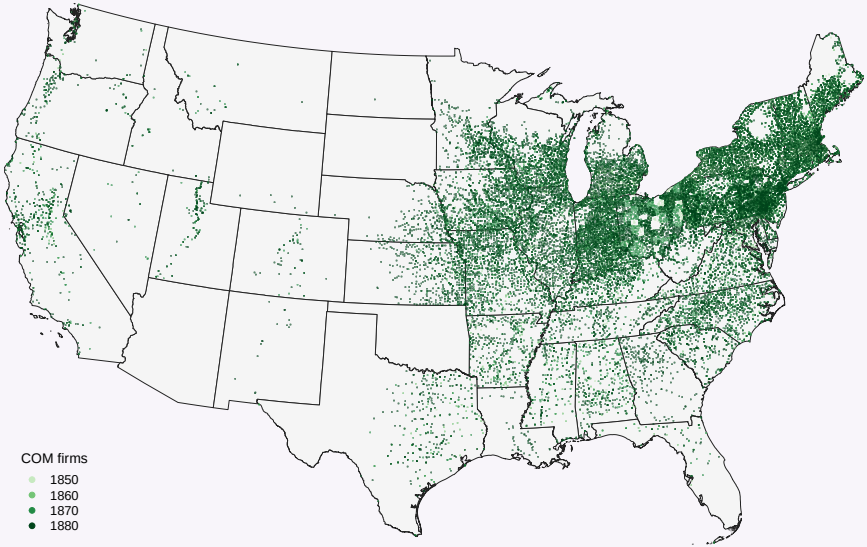
back

Appendix: 1880 COM Summary Statistics

	N	Mean	Median	Std. dev.	Min	Max
Most annual employees	202,419	12.22	4	55.79	0	9,000
Avg male emp (> 16)	195,280	7.71	2	33.30	0	6,000
Avg female emp (> 15)	20,696	15.97	3	63.21	0	3,000
Avg child emp	16,335	17.48	2	1,112.40	0	140,000
Avg day wages (skilled)	186,009	1.85	2	0.64	1	4
Avg day wages (unskilled)	155,202	1.05	1	0.41	0	3
Total annual wages	204,412	3,671.15	700	35,667.11	0	10238087

back

Appendix: COM Firms Map



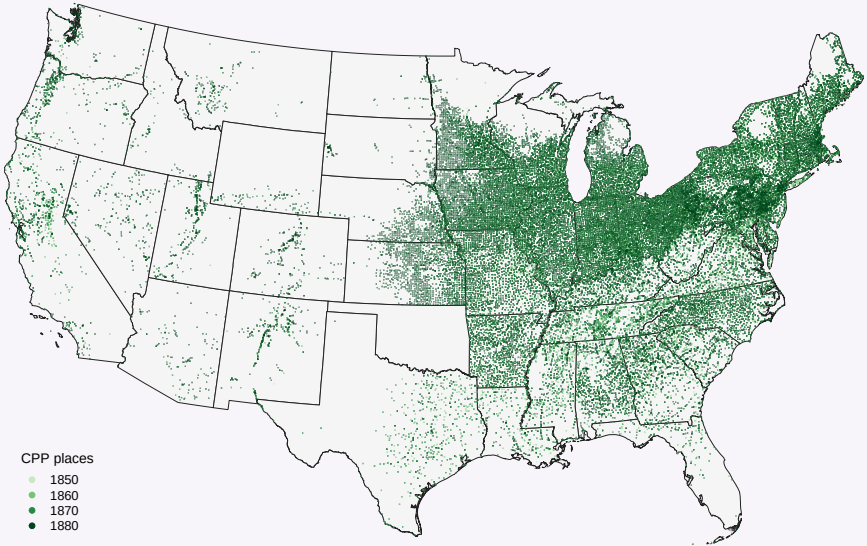
back

Appendix: COP Summary Statistics

	N	Mean	Median	Std. dev.	Min	Max
Male	16,671,732	0.84	1.0	0.37	0.0	1.0
Urban	16,671,732	0.32	0.0	0.47	0.0	1.0
White	16,671,732	0.85	1.0	0.35	0.0	1.0
Farm	16,671,732	0.40	0.0	0.49	0.0	1.0
Age	16,671,732	34.86	31.0	14.62	16.0	127.0
Literate	16,671,732	0.82	1.0	0.38	0.0	1.0
Occupational income score	16,671,732	18.71	14.0	10.27	3.0	80.0
Duncan socioeconomic index	16,671,732	19.66	14.0	18.38	3.0	96.0
Occ prestige score, Siegel	16,671,732	31.99	32.6	12.50	0.0	81.5
Occ earnings score	16,671,732	31.96	9.9	47.25	0.8	999.9
Occ education score	16,671,732	10.15	4.6	42.36	0.0	999.9
NPB occ stat score	16,671,732	31.57	18.9	45.75	0.7	999.9

back

Appendix: CPP Places Map



back

Appendix: COLS Income Distribution



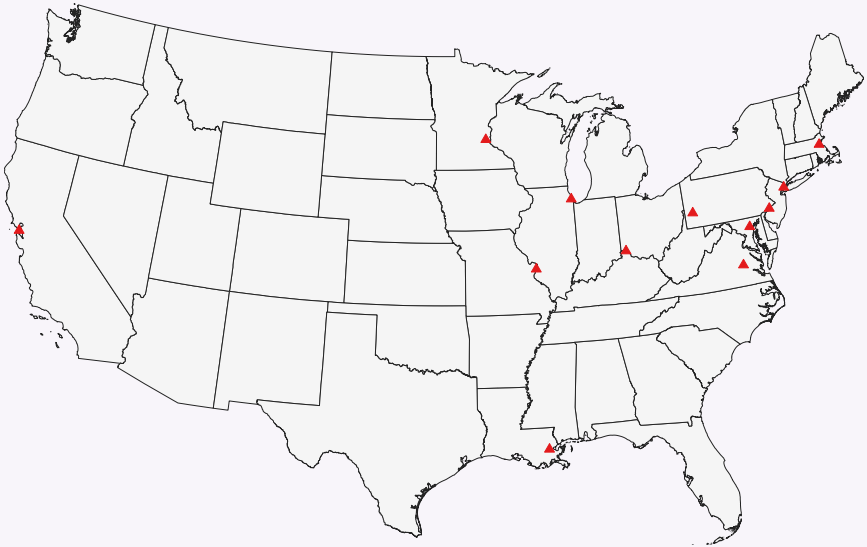
back

Appendix: COLS Summary Statistics

	N	Mean	Median	Std. dev.	Min	Max
Native-born HH head	6,627	0.54	1	0.50	0	1
Wife is working	6,536	0.07	0	0.25	0	1
# of Children	6,627	2.81	3	2.13	0	13
Family size	6,627	4.80	5	2.12	1	15
Husband age	6,627	39.43	38	11.01	18	99
Wife age	6,627	35.54	34	11.91	0	99
Husband's income	6,627	589.34	512	331.81	0	4,986
Wife's income	6,627	12.69	0	57.24	0	886
Children's income	6,627	107.49	0	230.84	0	1,989
Boarder income	6,627	35.98	0	107.82	0	1,828
Other income	6,627	15.57	0	62.15	0	1,330
Total income	6,627	761.07	665	372.44	93	4,986
Avg local income	6,627	591.61	625	130.98	305	1,074

back

Appendix: DOL Cities Map



back

Appendix: HLSP Summary Statistics

	N	Mean	Median	Std. dev.	Min	Max
Survey year	64,855	1894.42	1,894	4.55	1,879	1,903
Age in years	59,191	28.42	26	10.85	8	85
Skilled	64,782	0.20	0	0.40	0	1
Adult male	61,812	0.67	1	0.47	0	1
Adult female	61,812	0.29	0	0.45	0	1
Child	61,812	0.04	0	0.20	0	1
County FIPS coded	64,855	0.88	1	0.33	0	1
Daily wage, 1880 dollars	62,928	1.62	1	1.02	0	15
Weekly wage, 1880 dollars (6-day week)	62,930	9.74	8	6.14	0	88
Own annual earnings, 1880 dollars	64,250	438.44	372	302.42	1	4,556
Family annual earnings, 1880 dollars	7,063	456.74	432	255.71	2	3,385

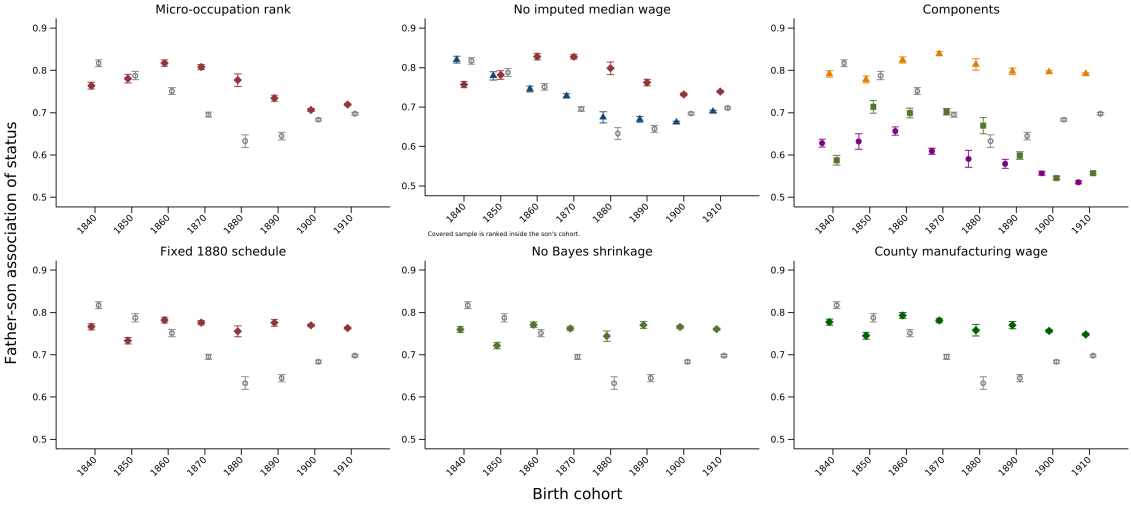
back

Appendix: Wage Data Coverage

Source	Workers	Share
Manufacturing, spatial	1,271,415	7.63%
Manufacturing, county fallback	641,938	3.85%
Manufacturing, state fallback	536,732	3.22%
Manufacturing, other fallback	124,199	0.74%
GPIH, state	10,580,489	63.46%
GPIH, region	231,522	1.39%
GPIH, national	170,183	1.02%
Uncovered	3,115,254	18.69%
Total	16,671,732	100.00%

back

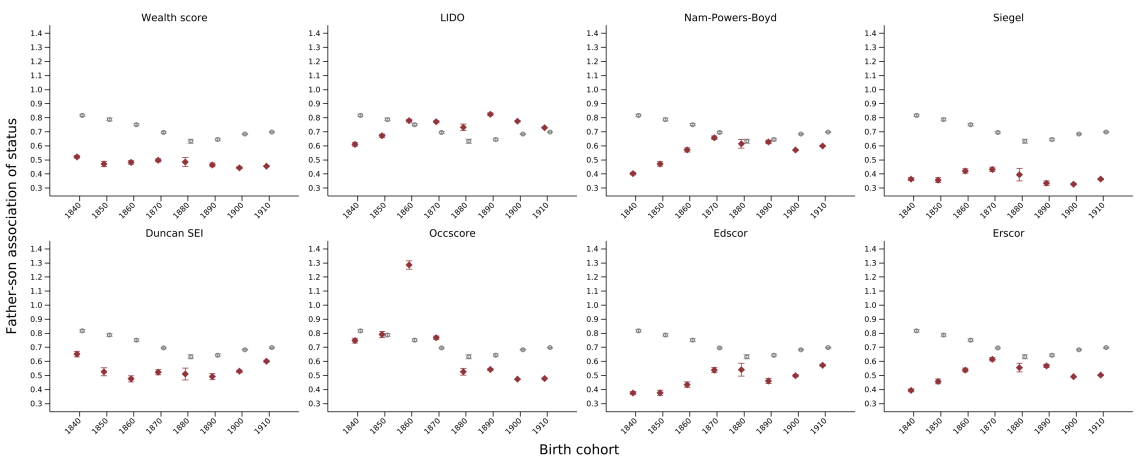
Appendix: Mobility Robustness



Micro-occupation mean, ranked among men aged 25-64. Slave fathers at the bottom.

back

Appendix: Mobility Alternative SES



Micro-occupation mean, ranked among men aged 25-64. Gray circles are the Ward replication.

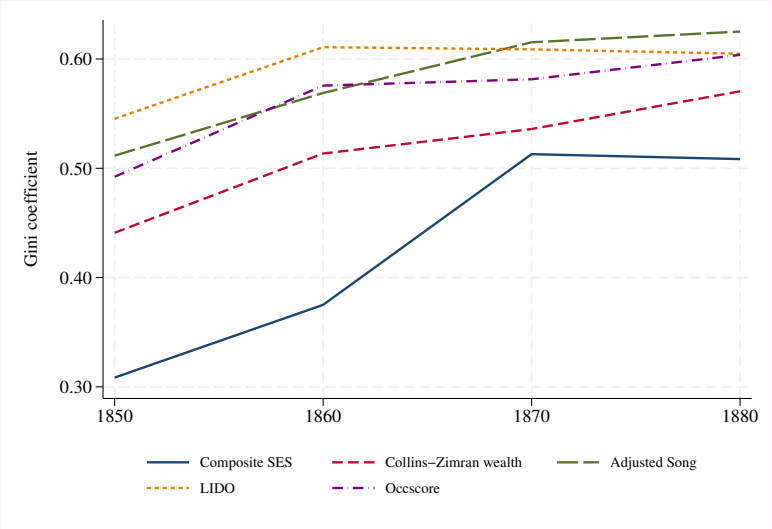
back

Appendix: Inequality by Components



back

Appendix: Inequality by SES



back

Appendix: All SES (All Worker p90/p10)

	1850	1860	1870	1880	1880−1850
Composite SES	2.19	2.81	17.07	13.14	10.95
Equal weights	2.32	4.74	20.27	13.44	11.11
Spatial wage	2.17	2.92	2.93	4.48	2.31
Average wealth	2.71	27.36	5.54	4.65	1.94
Siegel prestige	7.70	7.70	7.70	7.70	0.00
Occupational education	1.20	1.22	1.22	1.25	0.05
Collins-Zimran wealth	6.61	6.85	6.85	6.85	0.25
Duncan SEI	2.69	3.22	3.22	3.22	0.53
Occupational earnings	7.94	10.65	9.81	9.89	1.95
Occscore	3.69	9.22	7.10	7.10	3.41
Nam-Powers-Boyd	5.40	9.27	9.67	9.67	4.26
Adjusted Song	11.77	17.15	15.20	16.67	4.90
LIDO	4.21	12.70	9.80	10.13	5.92
Song	8.83	16.59	16.49	16.71	7.88

back