Unions and the Great Compression of Wage Inequality in the United States at Mid-Century
Supplemental Appendix

William J. Collins and Gregory T. Niemesh

August 2017

Preliminary: Please do not cite or circulate without authors' permission.

Contacts: william.collins@vanderbilt.edu; niemesgt@miamioh.edu.

Acknowledgements: The authors gratefully acknowledge the assistance of IPUMS-USA staff members in using the new 1960 5-percent sample, as well as helpful discussions with Ran Abramitzky, Leah Boustan, Brant Callaway, James Feigenbaum, Andrew Goodman-Bacon, Taylor Jaworski, Ilyana Kuziemko, Robert Margo, Suresh Naidu, Marianne Wanamaker, Ariell Zimran, and seminar participants at Queen's University (Canada), the National Bureau of Economic Research (NBER), and the World Congress of Cliometrics. Collins is the Terence E. Adderley Jr. Professor of Economics at Vanderbilt University and a Research Associate of the NBER. Niemesh is an Assistant Professor of Economics at Miami University and Faculty Research Fellow at the NBER.

A**1. Inequality Measures**

We make extensive use of decennial census microdata provided by IPUMS: the 1940 full count, the 5 percent extract for 1960, and the 1 percent extracts for 1950, and 1970-2000 (See Ruggles *et al*. 2015). Sample restrictions follow those in Goldin and Margo's (1992) study of the Great Compression: men aged 18-64, reporting at least 40 weeks of work as primarily a wage and salary worker, and an annual income at least half the minimum wage, on average, earned on a full-time basis for the reported number of weeks worked. In practice, this consists of limiting the sample to observations with calculated weekly earnings above the following yearly cutoffs:

	1939	1949	1959	1969	1979	1989	1999
Min wage per hour ($)	0.30	0.40	1.00	1.30	3.10	3.35	5.15
Per week at (1/2) full-time basis (40 hrs.) ($)	6	8	20	26	62	67	103

Weeks worked are reported in intervals in the 1960 and 1970 censuses. We impute weeks as the middle point of each interval in these years. Top-coded annual incomes are multiplied by 1.4 for 1940-1980. Top-coded values for 1990 and 2000 are set as the median or mean value of top-coded values in the respondent's state. Observations in the 1940 full count require the user to top-code wage and salary income values at $5,000. The instructions for the census enumerators state, "*For amounts above $5,000, enter '5,000+.' This means that you are not to report the actual amount of money wages and salary for persons who have received more than $5,000."* A number of enumerators recorded dollar amounts above $5,000 in error.

	1940	1950	1960	1970	1980	1990	2000
Top-code cutoff for annual income ($1,000)	5	10	25	50	75	140	175
Share of sample topcoded (%)	1.1	1.2	0.4	0.2	0.5	0.7	1.3

Observations are weighted by person weights (sample line weight for 1950). The sample includes the continental United States. In each year and location, we calculate the variance, gini, and 90-10, 90-50, and 50-10 percentile difference in log weekly wages.

A2. Geographic Identifiers

Recent literature on the post-1970s increase in inequality uses commuting zones (Autor and Dorn 2013; Autor, Dorn, and Hanson 2013) or metropolitan statistical areas (Moretti 2011; Baum-Snow and Pavan 2013) as the relevant geographic unit. We use a definition of local labor market that captures the geographic distribution of economic activity at mid-century, the time-period under consideration. While MSAs were defined for the 1940 Census, only a small number existed relative to today (140 in 1940 and 382 in 2010). Moreover, the changes in earnings inequality outside of metropolitan areas play an important role in the Great Compression that we would like to capture.

For these reasons, we use the State Economic Area (SEA) as our definition of a local labor market. The Census Bureau created SEAs to capture and tabulate data for single counties or contiguous groups of counties that were economically similar, and thus provide a similar notion to a local labor market, although different from Commuting Zones. In addition to economic characteristics, the Census Bureau used social, industrial, commercial, demographic, climatic, physiographic, and cultural factors into account when delaminating SEAs. SEAs are "*relatively homogenous subdivisions of States. They consist of similar economic and social characteristics. The boundaries of these areas have been drawn in such a way that each State is subdivided into a few parts, with each part having certain significant characteristics which distinguish it from the other areas which it adjoins*" (Bogue 1951).

The 1940 full count dataset contains county identifiers, which are easily aggregated into SEAs. The 1950 1 percent sample provides SEA identifiers as the smallest geographical unit. After 1950, neither SEA nor county identifiers are provided in the microdata samples. The smallest geographic unit in the 1960 sample is the mini public-use microdata area (mini-PUMA), created by combining census tracts and untracted counties to form areas of at least 50,000 persons or more.[1] The smallest geographic identifier in 1970 and 1980 is the county group (CNTYGP97 and CNTYGP98), each designed to contain at least 250,000 and 100,000 persons, respectively. The 1990 and 2000 samples use public-use microdata areas (PUMAs), again designed to contain at least 100,000 persons.

We construct SEA level indicators for 1960 and later samples following a procedure developed in Autor and Dorn (2013) by probabilistically matching county-groups and PUMAS in the census public use files to SEAs. The following description is heavily paraphrased from Dorn (2009).

[1] See https://usa.ipums.org/usa/volii/1960geotools.shtml for a detailed discussion of how mini-PUMAs were constructed by IPUMS.

Each MINIPUMA, PUMA, or CG j = 1,...,J is related to every SEA s = 1,...,467 by computing the probability that a resident of j lives in s using the following equation,

$$\alpha_{js} = \sum_{c=1}^{C} \frac{r_{jc}}{r_j} \frac{r_{cs}}{r_c}$$

where r_j is the number of residents in MINIPUMA j, r_c is the number of residents in county c, r_{jc} is the number of residents in the overlap between MINIPUMA j and county c, and r_{cs} is the number of residents in the overlap between county c and SEA s. The second term is either zero or one as each county is fully contained in a single SEA.

The overlap of MINIPUMA j (or PUMA or CG) and county c (r_{jc}) is unobservable in the microdata. For 1960, IPUMS provides a crosswalk of MINIPUMA population in each county, which can be used in conjunction with total population counts for counties from Haines (2010) to compute r_{jc}. The alpha factor weight would be equal to 1 for MINIPUMAs contained within a single SEA, even if spread over multiple counties. The alpha factor weight on a single census observation will be strictly between 0 and 1 when a MINIPUMA spans multiple SEAs. For example, suppose 30 percent of MINIPUMA j's population resides in SEA 1, and 70 percent resides in SEA 2. Then each observation in the MINIPUMA j will have two copies in the dataset, one in SEA 1 a weighting factor of $\alpha_{j1} = 0.30$, and the other with $\alpha_{j2} = 0.70$. A similar procedure is completed for 1970 through 2000 where the MINIPUMA is replaced by the CG or PUMA.

A3. Union Density Rates

Troy (1957) provides industry and state-level union density measures for 1939 and 1953. We construct local measures of industry structure using the full-count 1940 census microdata provided by IPUMS. The local change-in-union exposure variable for SEA s (ΔU_s), is a weighted average of national industry-level changes in unionization as a percent of employment for each industry j (ΔU_j), where the weights (ω_{js}) correspond to the local mix of employment in 1940 measured as the employment in industry j in SEA s divided by total employment in SEA s $\omega_{js} = \frac{E_{js}}{E_s}$, where employment is based on the same sample restrictions as used to make the inequality measures.

$$\Delta U_s = \sum_{j=1}^{N} \Delta U_j \times \omega_{js}$$

Table A1 reports how we applied union density rates by industry reported in Troy (1957) to the 1950 industry codes used in the 1940 full count census.

A4. Construction of labor demand control variables

We construct two labor demand indexes from IPUMS census microdata. The first is a Bartik (1991) style index using initial local-level industry mix in 1940 and national-level industry growth rates. The second demand index allows for skill-specific changes in demand at the local level following the index used in Goldin and Margo (1992).

Let *j* denote industry, *s* denote SEA, and *t* denote year. The first demand index (non-skill specific) is constructed according to the following equation:

$$D_s^{Bartik} = \sum_j \left(\frac{E_{js1940}}{E_{s1940}}\right)\left(\frac{E_{j1950} - E_{j1940}}{E_{j1940}}\right)$$

For a given SEA, the first term captures the percent of total employment made up by industry *j*, while the second term captures the national growth rate in employment for industry *j*. SEAs in which fast-growing industries (nationally) make up a larger share of employment will have a higher value for this index.

The second index captures the fact that the skill-mix varies by industry and within an industry *j* the skill-mix varies across SEAs. Thus, the relative demand for skill deriving from employment changes in that industry would vary across SEAs for a given industry growth rate, causing differential impacts on the local wage distribution. Following Goldin and Margo (1992), we let *i* denote skill category, and can find the local demand for skill group *i* in year *t* by,

$$D_{ist} = \sum_j \left(\frac{E_{jt}}{E_t}\right)\left(\frac{E_{ijs40}}{E_{js40}}\right)\left(\frac{E_{js40}}{E_{s40}}\right)$$

The first term captures the national employment share of each industry. Varying over time as industries wax and wane, it provides the sole source for changes in the index over time within an SEA. The second term is fixed at 1940 base levels and captures the SEA-specific skill mix within each industry. The third term captures the industry mix at the SEA level, and is fixed at 1940 base levels. The index can be simplified by canceling out the denominator in the second term and the numerator in the third term to get.

$$D_{ist} = \sum_j \left(\frac{E_{jt}}{E_t}\right)\left(\frac{E_{ijs40}}{E_{s40}}\right)$$

The index above is skill specific for each SEA. To get a relative skill demand index for year *t*, we use the following ratio of skilled to unskilled,

$$D_{s,t}^{GM} = \frac{D_{(skilled)st}}{D_{(unskilled)st}} = \frac{\sum_j \left(\frac{E_{jt}}{E_t}\right)\left(\frac{E_{(skilled)js40}}{E_{s40}}\right)}{\sum_j \left(\frac{E_{jt}}{E_t}\right)\left(\frac{E_{(unskilled)js40}}{E_{s40}}\right)}$$

Finally, the decadal change in local relative skill demand is found by taking $D_{s,1950}^{GM} - D_{s,1940}^{GM}$.

A5. Construction or sources of other control variables

Using IPUMS samples (Ruggles et al. 2015):

- Share of males that are high school graduates in 1940, 1950, 1960: percent of males fitting sample restrictions for inequality measures that reported completing four years of high school.
- Share of wage and salary workers earning below the minimum wage in 1939: percent of males aged 18-64, earning positive wage and salary income, reporting wage and salary employment as their primary work, that earn less than the minimum wage on a weekly full-time basis (or less than $12 weekly in 1939 - $0.30 per hour times 40 hours).
- Variables measuring changes in inequality in occupational standing from 1920 to 1940: We use two measures of occupational standing. The first, provided by IPUMS, is the standard *occscore* variable, which "assigns each occupation in all years a value representing the median total income (in hundreds of 1950 dollars) of all persons with that particular occupation in 1950." Using the 1950 income distribution bakes in the large cross-occupation compression in wages of the "Great Compression." We create a second and more flexible occupational standing variable from the 1940 income distribution of male workers that satisfy the sample restrictions used in the main inequality measures. For observations in industry-by-occupation cells with more than 50 observations, we create an *occscore40* variable that assigns the median weekly earnings of that cell. For cells with 50 observations or less, the national median weekly earnings for the occupation is assigned. Each of the occscore measures are then assigned to observations in 1920, 1930, and 1940 based on the occupational code. We calculate inequality measures from 1920-40 and 1930-40: change in log 90-10, change in log 90-50, change in log 50-10, change in the Gini of the log occscore, and the change in the variance of log occscores.

Calculated from Haines (2010)

- The male employment rate in 1940 (the percent of males aged 14+ that are employed)
- Percent of total population that resides in an urban area in 1940
- Per capita WWII contracts in dollars

BLS Cost of Living Report 1917-1919 (ICPSR 8299)

- City-level inequality in 1919 for a reduced sample of 83 cities: In combination with the 1940 full count IPUMS data, we estimate the change in local wage inequality from 1919 to 1939. The BLS survey provides micro-level income data for a sample of more than 12,000 families in 99 cities. Unfortunately for the purposes of this study, the sample frame was narrowed to focus on urban middle-class families. The BLS limited the sample to husband-wife families with at least one child were surveyed, families not living in "slums", and to families earning less than $2,000 a year (though they were not always). Families were to have resided in the place for at least one year, were not to have more than three boarders or lodgers, and were not to be non-English speakers who had resided in the US for less than five years. Feigenbaum (2016) estimates that the BLS sample is concentrated between the 30th to 70th percentiles of the national income distribution. Note that the data were made available (in part) by the Inter-university Consortium for Political and Social Research (ICPSR). The data for *Cost of Living in the United States, 1917-1919* were originally collected by the Bureau of Labor Statistics. Neither the collector of the original data nor the consortium bear any responsibility for the analyses or interpretations presented here.

Other sources:

- State mobilization rate in WWII: Acemoglu, Autor, and Lyle (2004).
- Right to work state in 1950: Indicator variable for whether the state had passed "right to work" legislation by 1950. Source: Ellwood and Fine (1987, Table 1).

A6: Tables

Table A1: Crosswalk of industry categories reported in Troy (1957) to 1950 industry codes in IPUMS census microdata

Troy (1957) industry categories	1950 industry codes
Manufacturing	
Metals	336 - 399
Clothing	448 - 449
Food, liquor, and tobacco	406 - 429
Paper, printing and publishing	456 - 459
Leather and leather products	487 - 489
Chemicals, rubber, clay, glass, and stone	316 - 328, 466 - 478
Textiles	436 - 446
Lumber and lumber products	306 - 309
Transportation (exclusive of railways), communications, and public utilities	507 - 598
Railway transportation	506
Building and construction	246
Mining, quarrying, and oil	206 - 239
Public service	906 - 936
Services	606 - 899
Apply the manufacturing average to “non specified manufacturing”	499

Table A2: Full estimation results for base specification (Row 1 of Table 2)

Panel A: 1940-50	(1)	(2)	(3)	(4)	(5)
	Δ90-10	Δ90-50	Δ50-10	ΔGini	ΔVar
Change in union exposure	-0.0232	-0.0030	-0.0201	-0.0041	-0.0066
	(0.0062)	(0.0033)	(0.0044)	(0.0011)	(0.0019)
Median log weekly wage 1939	-0.2622	0.2673	-0.5294	0.0150	-0.0544
	(0.0515)	(0.0304)	(0.0545)	(0.0120)	(0.0222)
Initial log 90-10 gap in 1939	-0.5518	-0.2703	-0.2817	-0.1043	-0.2032
	(0.0657)	(0.0366)	(0.0585)	(0.0113)	(0.0212)
Per capita WWII contracts (1940$)	0.0024	-0.0021	0.0044	-0.0002	-0.0009
	(0.0052)	(0.0026)	(0.0052)	(0.0009)	(0.0017)
Male employment rate in 1940	0.6164	0.0001	0.6161	0.1626	0.2754
	(0.1991)	(0.0999)	(0.1579)	(0.0434)	(0.0800)
Percent urban 1940	0.0335	-0.0869	0.1202	-0.0017	-0.0250
	(0.0479)	(0.0328)	(0.0459)	(0.0094)	(0.0183)
Percent males HS grads	-0.5056	-0.0397	-0.4665	-0.1060	-0.1465
	(0.1277)	(0.0970)	(0.1397)	(0.0448)	(0.0669)
Local demand shock index	0.0951	0.0163	0.0785	0.0219	0.0322
	(0.0613)	(0.0270)	(0.0462)	(0.0087)	(0.0170)
Skill-specific demand shock index	1.4103	0.7764	0.6330	0.1864	0.3137
	(0.4659)	(0.2344)	(0.3746)	(0.0734)	(0.1472)
Midwest census region	-0.0081	0.0069	-0.0148	0.0016	0.0057
	(0.0143)	(0.0120)	(0.0165)	(0.0050)	(0.0064)
South census region	0.0391	0.0334	0.0057	0.0069	0.0017
	(0.0218)	(0.0165)	(0.0170)	(0.0045)	(0.0063)
West census region	0.0466	-0.0131	0.0598	0.0065	0.0170
	(0.0234)	(0.0136)	(0.0238)	(0.0062)	(0.0092)
Constant	1.0296	-0.5783	1.6084	-0.0463	0.2555
	(0.2387)	(0.1373)	(0.2236)	(0.0520)	(0.0950)
Observations	467	467	467	467	467
R-squared	0.3215	0.4483	0.5004	0.2859	0.2970

Continued Table A2: Full estimation results for base specification (Row 1 of Table 2)

Panel B: 1940-60	(1)	(2)	(3)	(4)	(5)
	Δ90-10	Δ90-50	Δ50-10	ΔGini	ΔVar
Change in union exposure	-0.0214	-0.0048	-0.0165	-0.0032	-0.0056
	(0.0047)	(0.0025)	(0.0046)	(0.0008)	(0.0015)
Median log weekly wage 1939	-0.1727	0.3570	-0.5298	0.0346	-0.0266
	(0.0496)	(0.0410)	(0.0431)	(0.0107)	(0.0170)
Initial log 90-10 gap in 1939	-0.6024	-0.3047	-0.2978	-0.1252	-0.2341
	(0.0546)	(0.0420)	(0.0601)	(0.0109)	(0.0177)
Per capita WWII contracts (1940$)	0.0135	0.0050	0.0085	-0.0001	0.0016
	(0.0059)	(0.0025)	(0.0042)	(0.0011)	(0.0016)
Male employment rate in 1940	0.3368	0.1853	0.1509	0.0922	0.0761
	(0.1035)	(0.0599)	(0.1128)	(0.0241)	(0.0346)
Percent urban 1940	0.1788	-0.0824	0.2609	0.0100	0.0343
	(0.0324)	(0.0223)	(0.0295)	(0.0083)	(0.0127)
Percent males HS grads	-0.3675	-0.1084	-0.2593	-0.0880	-0.1296
	(0.1516)	(0.1106)	(0.1551)	(0.0290)	(0.0520)
Local demand shock index	0.1142	0.0094	0.1045	0.0191	0.0464
	(0.0411)	(0.0127)	(0.0438)	(0.0052)	(0.0106)
Skill-specific demand shock index	0.7190	0.5337	0.1872	0.1898	0.2833
	(0.4893)	(0.2148)	(0.4604)	(0.0649)	(0.1160)
Midwest census region	-0.0583	-0.0226	-0.0356	-0.0001	-0.0017
	(0.0136)	(0.0091)	(0.0113)	(0.0051)	(0.0062)
South census region	0.0688	0.0177	0.0513	0.0213	0.0312
	(0.0198)	(0.0123)	(0.0206)	(0.0044)	(0.0064)
West census region	0.0819	-0.0150	0.0971	0.0123	0.0388
	(0.0290)	(0.0109)	(0.0260)	(0.0057)	(0.0109)
Constant	0.8436	-0.8823	1.7270	-0.0419	0.2513
	(0.2311)	(0.1333)	(0.1958)	(0.0386)	(0.0673)
Observations	467	467	467	467	467
R-squared	0.5677	0.7922	0.6091	0.6433	0.6587

Notes: Each column represents the results from a separate regression and corresponds to the full estimation output of the base specification in row (1) of Table (2) in the main text. Region effects are relative to the Northeast census region.

Sources: Inequality measures, wage measures, demand indices, high school educational attainment, and are derived from the 1940, 1950, or 1960 IPUMS samples (Ruggles et al. 2015), as described in the text. Exposure to unions (ΔU_{ir}) is calculated by interacting industrial employment data from the 1940 full count census and industry-level changes in union density from Troy (1957). The employment rate and share of population in urban areas are calculated with data from Haines (2010).

Table A3: Changes in wage inequality and unionization, robustness to sample definitions

	(1) Δ90-10	(2) Δ90-50	(3) Δ50-10	(4) ΔGini	(5) ΔVar
Panel A: 1940-50					
(1) Base	-0.0232	-0.0030	-0.0201	-0.0041	-0.0066
	(0.0062)	(0.0033)	(0.0044)	(0.0011)	(0.0019)
(2) Drop Ag workers	-0.0155	-0.0020	-0.0134	-0.0028	-0.0033
	(0.0051)	(0.0032)	(0.0040)	(0.0010)	(0.0016)
(3) Δ1919-39 in inequality	0.0072	-0.0194	0.0266		-0.0112
(83 cities)	(0.0219)	(0.0168)	(0.0188)		(0.0096)
Panel B: 1940-60					
(1) Base	-0.0214	-0.0048	-0.0165	-0.0032	-0.0056
	(0.0047)	(0.0025)	(0.0046)	(0.0008)	(0.0015)
(2) Drop Ag workers	-0.0131	-0.0036	-0.0094	-0.0019	-0.0027
	(0.0049)	(0.0025)	(0.0049)	(0.0008)	(0.0015)

Notes: Each figure in the table is the estimated coefficient (τ) on the change-in-unions variable from a separate regression of the base specification on different samples. The base specification's control variables are described in the text; they include regional fixed effects, measures of the 1939 wage structure (the median log wage and the 90-10 wage gap), 1940 economic conditions (the employment rate for men, the share of workers who completed high school, and the percent of population in urban areas), wartime demand and investment shocks as reflected in war production and facilities contracts per capita, and labor demand shift variables. Row (1) reports results from the base specification from Table 2. Row (2) drops all wage and salary farmers, farm managers, and farm laborers. The dependent variable in row (3) is the change in inequality from 1919-1939 for a set of 83 cities in the BLS 1919 Cost of Living Survey.
Sources: See notes to table A2. We derive 1919 inequality from the Bureau of Labor Statistics 1919 Cost of Living Survey (ICPSR Study 8299).

Table A4: Effect of union exposure on (residual) within-group inequality

	(1) $d\sigma_{5040}$	(2) $d\sigma_{5040}$	(3) $d\sigma_{6040}$	(4) $d\sigma_{5040}$
Change in union exposure	-0.0015	-0.0022	-0.0020	-0.0029
	(0.0010)	(0.0009)	(0.0008)	(0.0008)
$\sigma_{s,40}$		-0.5542		-0.7368
		(0.1315)		(0.0726)
Median log weekly wage 1939	-0.0350	-0.0190	-0.0225	-0.0013
	(0.0151)	(0.0155)	(0.0108)	(0.0095)
Initial log 90-10 gap in 1939	-0.1069	-0.0109	-0.1232	0.0044
	(0.0152)	(0.0234)	(0.0103)	(0.0162)
Per capita WWII contracts (1940$)	-0.0016	-0.0013	-0.0001	0.0002
	(0.0011)	(0.0011)	(0.0010)	(0.0009)
Male employment rate in 1940	0.2024	0.1729	0.0804	0.0412
	(0.0528)	(0.0483)	(0.0248)	(0.0179)
Percent urban 1940	0.0061	0.0104	0.0288	0.0346
	(0.0113)	(0.0100)	(0.0077)	(0.0072)
Percent males HS grads	-0.0790	-0.0653	-0.0974	-0.0792
	(0.0578)	(0.0594)	(0.0309)	(0.0280)
Local demand shock index	0.0053	-0.0066	0.0180	0.0021
	(0.0126)	(0.0129)	(0.0070)	(0.0058)
Skill-specific demand shock index	0.2438	0.1361	0.2068	0.0636
	(0.0946)	(0.0899)	(0.0687)	(0.0535)
Midwest census region	0.0043	0.0022	0.0026	-0.0002
	(0.0054)	(0.0049)	(0.0037)	(0.0022)
South census region	0.0092	0.0059	0.0225	0.0182
	(0.0050)	(0.0053)	(0.0046)	(0.0033)
West census region	0.0139	0.0044	0.0244	0.0118
	(0.0093)	(0.0091)	(0.0067)	(0.0043)
Constant	0.0962	0.0773	0.1185	0.0934
	(0.0694)	(0.0686)	(0.0384)	(0.0412)
Observations	467	467	467	467
R-squared	0.2064	0.2388	0.6077	0.7246
St. dev of dependent variable	0.0499		0.0271	
St. dev of change in union exp.	3.14			

Notes: Each column reports the full estimation results from a separate regression at the SEA-level. The dependent variable is the change in the variance of the residuals from a wage regression individually estimated for each SEA-by-year. The variance of predicted residuals for SEA *s* in year *t* ($\sigma_{s,t}$) is derived from the predicted residuals from a regression of log weekly earnings on experience, experience squared, years of schooling, years of schooling squared, the interaction of experience and years of schooling and an indicator for marital status (experience = age – years of schooling – 6). Columns (2) and (4) add the initial 1940 level of variance of the residuals. Region effects are relative to the Northeast census region.
Sources: See notes to table A2.

References

Acemoglu, Daron, David H. Autor, and David Lyle. 2004. “Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury.” *Journal of Political Economy* 112, 3: 497-551.

Autor, David H. and David Dorn, 2013. “The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market,” *American Economic Review* 103(5): 1553-97.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. “The China Syndrome: Local Labor Market Effects of Import Competition in the United States,” *American Economic Review* 103(6): 2121-68.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Baum-Snow, Nathaniel and Ronni Pavan. 2013. “Inequality and City Size,” *Review of Economics and Statistics* 95(5): 1535-1548.

Bogue, Donald. 1951. “State Economic Areas.” Department of Commerce, Bureau of the Census. Washington, D.C.

Dorn, David. 2009. “Essays on Inequality, Spatial Interaction, and the Demand for Skills.” Dissertation University of St. Gallen no.3613, September. Data Appendix, pp. 121-138.

Ellwood, David T. and Glenn A. Fine. 1987. “The Impact of Right to Work Laws on Union Organizing.” *Journal of Political Economy* 95, 2: 250-273.

Feigenbaum, James J. 2015. “Intergenerational Mobility during the Great Depression.” Unpublished working paper.

Goldin, Claudia and Robert A. Margo. 1992. “The Great Compression: The Wage Structure in the United States at Mid-Century.” *Quarterly Journal of Economics* 107, 1: 1-34.

Haines, Michael R., and Inter-university Consortium for Political and Social Research. 2010. *Historical, Demographic, Economic, and Social Data: The United States, 1790-2002* [computer file]. ICPSR02896-v3. Ann Arbor, MI: ICPSR [distributor].

Moretti, Enrico. “2011. Local Labor Markets.” *Handbook of Labor Economics,* vol 4, eds. Orley Ashenfelter and David Card. Elsevier.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. *Integrated Public Use Microdata Series: Version 6.0* [Machine-readable database]. Minneapolis: University of Minnesota.

Troy, Leo. 1957. *The Distribution of Union Membership among the States, 1939 and 1953*. New York: National Bureau of Economic Research.

United States Department of Labor. Bureau of Labor Statistics. Cost of Living in the United States, 1917-1919. ICPSR08299-v5. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1986. https://doi.org/10.3886/ICPSR08299.v5