Migration and Occupational Mobility in the United States, 1850-1940

Matthew Hurt

January 14, 2022

1 Introduction

The degree to which labor markets are well integrated across space and occupation have important welfare and distributional consequences for workers [1]. This is especially true in the United States, where the economy having high occupational and geographic mobility are essential elements of the American story (Ferrie 2005). One version of this story is that a national labor market emerged in the United States after the Civil War. How this integration unfolded unevenly matters in understanding the evolution of different parts of the American economy [2].

Until recently economic historians have had to rely on uneven data and approaches to measure labor market integration (Kim 98). Common approaches involved comparing aggregate population and occupation trends over time or relied on myopic wage and price data to make inferences about larger regions (Rosenbloom 90). Today the availability of high quality US census data at the individual level provides economic historians with new opportunities to better measure labor market integration (Abramitzky et al 2017).

This paper uses a discrete version of the dynamic approach employed in Artuc and McLaren (2015) to measure switching costs faced by workers when moving between different locations or occupations. As labor markets become better integrated these costs should approach zero. Studying the period between 1850 to 1880 and 1900 to 1940 decade by decade shows that the fixed costs of changing location or occupation increased between 1850 to 1880 and declined during the 20th century. Though direct comparisons of the two costs are not possible, the fixed costs of changing location fell roughly 15 percent while fixed costs of changing occupation fell roughly 25 percent.

Even when a labor market is well integrated, the welfare gains depend on the details of this integration. One can imagine a labor market that may be well integrated because

[1] Bryan and Morten (2019) show this for modern day Indonesia and use the modern United States as a reference.

[2] Rosenbloom (2002) documents a number of differences between regional labor markets in the United States over time with a mix of quantitative and qualitative evidence

low skilled workers face low switching costs when moving between low skilled occupations or across locations, but still face large hurdles when transitioning out of low skilled work. To better measure occupational integration in the United States additional costs for workers in low skilled occupation groups face when transitioning into the high skilled occupation group are also estimated. These costs also decline through time, except for farmers whose costs of switching into high skilled labor increase with time. In particular, the costs faced by unskilled workers transitioning into high skilled work declined nearly 45 percent while the costs faced by farmers increased by 8 percent.

This paper contributes to three strands of literature. The first is the literature already discussed above that focuses on quantifying labor market integration in the US during this period. This literature (Wright 1986,1987, Rosenbloom 1990, Collins and Wannamaker 2015, and Mitchener and McLean 1998) highlight two commonly observed trends. The first is that labor market integration increased during this period. The second is that Southern labor markets were uniquely different from the rest of the country. The evidence supports the first point and potentially the second point since farmers faced higher costs and the South had a greater share of workers in agriculture than other parts of the country.

The second literature is one that includes the intersection of trade theory and American economic history. For example, Kim (1998) and Steinwender (2018) have taken trade theory to historical data to improve our understanding of important economic forces at the time. In addition Artuc et al (2010) have generated models that are uniquely suited to constraints of this historical setting. This paper provides another example of how these rich models can be combined with historical data to continue improving our understanding of the time period.

The final literature this paper contributes to is one studying the effect of changes in migration patterns in the US. Abramitzky and Boustan (2017) summarize some of their and much of the other work done in this area, often using linked census data to measure immigrant outcomes. Ferrie (2005) and Salisbury (2014) focus on the effect of differential changes in migration rates across space, in particular declining rates of migration across state lines. Caselli and Coleman (2001) and Alston and Hatton (1991) highlight the impacts of the changing composition of the labor force during this period as agriculture lost its majority share of the workforce. This paper contributes to this literature as well by showing that these declining migration rates in conjunction with declining costs provide evidence of increasing labor market integration. This paper also shows that farmers faced statistically different costs compared to other groups when transitioning into high skilled work.

2 Empirical Methodology

To measure labor market frictions between 1850-1940 for different occupation groups using only information on changes in locations and occupations over time this paper employs a model similar to the one used by Artuc and McLaren (2015). In their paper they look at more recent US data on occupational and sectoral changes in employment. The approach is robust enough to use location instead of sectors as the second dimension of variation. The model involves using a two stage estimation strategy to first use the flow of workers between occupations and locations to estimate switching costs up to a preference parameter ν_t and then use information on wages and occupation-location cell attractiveness to estimate the costs of switching between cells. These costs can also be thought of as the willingness to pay of a worker to switch between cells. These switching costs are inversely related to labor market integration.

To motivate the framework consider a world in which workers begin each period in an endowed (origin) location and occupation cell. These workers receive perfect information about the wage and amenity values of all possible (destination) location and occupation combinations. Workers can choose to change their occupation, location, or both paying fixed costs to do so. Workers in any location-occupation cell receive the same wage and amenity. Workers always migrate to the occupation-location cell that maximizes utility net of any switching costs.

There are two forces that prevent workers from sorting into the location-occupation cell with the highest utility payout through wages and amenities. The first force is due to the initial endowment workers receive. If costs are sufficiently high, workers will stay in their original cell. The second force is that all workers draw a worker specific shock from an extreme value distribution with variance parameter ν_t. When these shocks are large, workers prioritize cell amenity value over cell wages. Empirically, we expect ν_t to decline over time as wages become more valuable relative to amenities due to increases in good varieties and quality. Based on this framework a two stage approach is used to separate out push and pull forces of migration along with the value of ν_t for each year pair.

Artuc and McLaren show that in the environment described above one can use an estimation equation like (1) to estimate origin and destination cell fixed effects and switching costs up to the parameter ν_t

$$Z_t^{ijkl} = exp(\alpha_t^{ik} + \lambda_t^{jl} + \frac{C_t(i,j,k,l)}{\nu_t}) + \varepsilon_t^{ijkl} \tag{1}$$

In the first stage regression, i, j are the origin and destination location choices while k, l are the origin and destination occupation choices. Here t represents two linked years. The linking procedure is explained in the data section. The equation has the structure of a gravity model with α_t^{ik} and λ_t^{jl} functioning as the origin location-occupation and destination location-occupation fixed effects. The bilateral frictions are measured by $\frac{C_t(i,j,k,l)}{\nu_t}$. Z_t^{ijkl} is the number of people in each year of each possible combination of the four dimensions. Zeroes are included. For estimation purposes λ_t^{11} is the excluded reference group.

This equation is flexible with respect to how the cost function is constructed. For simplicity in estimation and interpretation of results the cost function used is constructed using different fixed costs. The first component of the cost function takes on a value of 1 if workers change their location. The second component takes on a value of 1 if workers change their occupation. The third component takes on a value of 1 if workers change both. The most complicated piece is the last set of costs which take a value 1 if the worker's origin occupation

k is not in the high skilled occupation group, but their destination occupation l is in the high skilled occupation group.

$$C_t(i,j,k,l) = \beta_{1t}1_{i\neq j} + \beta_{2t}1_{k\neq l} + \beta_{3t}1_{both} + \gamma_{t,k\neq 1,l=1} \tag{2}$$

β_{1t} measures the fixed cost workers pay when they change location, regardless of whether they change occupation. Migration will increase as this cost falls because marginal workers who can secure higher utility in a different cell than their origin cell can now move. Empirically we expect this to decline as the real costs of moving across space decline. β_{2t} measures the fixed cost workers pay when they change occupation, regardless of whether they change location. Migration will increase as this cost decreases because marginal workers can now secure higher utility by changing occupations. β_{3t} measures the additional cost workers who change location and change occupation face. Unlike the first two costs the sign on this coefficient may not be negative. Workers may have to pay an additional cost if coordinating changes between locations and occupations is expensive. However, if workers can bundle costs β_{3t} may be positive.

The $\gamma_{t,k\neq 1,l=1}$ measure the additional fixed cost paid by members of each occupation group when transitioning into high skilled work. The variation used to estimate these terms is slightly different than the ones already measured. Each worker in these groups are already paying β_{1t} since they are changing occupation. However, β_{1t} is capturing occupation changes not captured by the $\gamma_{t,k\neq 1,l=1}$. Workers in this group are also paying β_{2t} and β_{3t} if they change location while moving from a low skilled to the high skilled occupation group. The data section highlights the shift in employment into white collar work taking place during this time. These costs provide new insight into how this transition took place.

Equation (1) is estimated separately for each of the decadal year pairs 1850-1860, 1860-1870, 1870-1880, 1900-1910, 1910-1920, 1920-1930, and 1930-1940. Given the non-linear nature of the right hand side expression of equation (1) the authors guidance is followed and use pseudo-maximum likelihood (PPML) estimation is used for the regression strategy.

PPML has attractive theoretical properties such as estimating the equation without a biased error term. [3]. It also has the advantage of rationalizing empty cells on the left hand side. In migration an empty cell communicates important information about push and pull forces between any empty cells and all others. Empty cells are common in the data so a technique that incorporates them is necessary for estimation.

For each of the year pairs ν_t is estimated in the second stage by regressing a combination of the origin and destination fixed effects on the wage in the destination cell. Since the fixed effects incorporate push and pull factors related to amenity values and wages once wages are controlled for it is possible to determine the importance of amenity values in the migration decision. Equation (3) is a simplified version of the second stage used by Artuc and McLaren.

$$\alpha_t^{ik} + \lambda_t^{ik} = \frac{1}{\nu_t} w_{t+1}^{ik} + \epsilon_t^{ik} \tag{3}$$

Here the notation is slightly different. Since workers can have any cell as their origin and any cell as their destination, each location-occupation cell has a corresponding origin and destination fixed effect. In this regression, $\alpha_t^{ik}, \lambda_t^{ik}$ are the fixed effects for each ik location-occupation cell. w_{t+1}^{ik} is the wage in the destination (second) year for each cell. Since workers are endowed with their origin cell and choose their destination cell we only need the destination wage for this estimation strategy. After estimating ν_t it is possible to back into the true $C_t(i, j, k, l)$.

3 Data

The data on worker location and occupation are available for each decade between 1850 and 1940, excluding 1890 and come from IPUMS (Ruggles et al. 2019). The data that link individuals between each censuses currently exist for each pair of those decades except 1890 and come from the Census Linking Project (Abramitzky et al 2020). From these two sources

[3] Silva & Tenreyro (2006) demonstrate the particular benefits of using PPML when estimating gravity equations. For greater efficiency in estimation I use the command developed by Correia et al.

six sets of year pairs from 1850 to 1860, 1860 to 1870, 1870 to 1880, 1900 to 1910, 1910 to 1920, 1920 to 1930, and 1930 to 1940 are used. Equations (1) and (3) are ran separately for each of the year pairs.

The data only include male workers who held a known occupation in the origin and destination year, lived in the continental US in the origin and destination year, and were age 16 to 55 in the origin year and age 26 to 65 in the destination year. The locations include the 48 continental United States and the District of Columbia. For clarity in estimation occupations are condensed into five different occupation groups based on the workers written description of their occupation[4]. The white collar occupation group consists of highly skilled workers.

Since there are 49 locations and five occupations each year has a total of 245 possible cells. Once two years are linked there are 60025 location occupation origin destination cell pairs. Next the number of workers in each of these 60,025 cells are recorded. Cells where there are no workers receive a zero value. For the 1850 to 1860, 1860 to 1870, and 1870 to 1880 year pair data fewer than the total possible 60,025 cell pairs are used since some states did not exist by then. Instead all available cells are used in each year. Concerns about how changing geography may be driving results in the pre 20th century data are valid, but sub-sample analysis of the 19th and 20th century data reveal similar trends.

Using these data we can already measure several key trends in our data that help us contextualize the costs we later measure. First, the major occupational change taking place in the US between 1850 to 1940 is the transition of workers away from agriculture and into other forms of labor, particularly white collar work. Figure 1 is created using data from the unlinked full count census data with a similar restriction the one described above except that all male workers age 16 to 65 are included. The figure shows that nearly half of all workers were engaged in the agricultural work as late as 1870. By 1940 barely 20 percent of workers were.

[4]Collins & Zimran (2019) utilize a classification scheme I make only one deviation to by grouping farm laborers in the *farmer* occupation group rather than including them in the *unskilled* group.

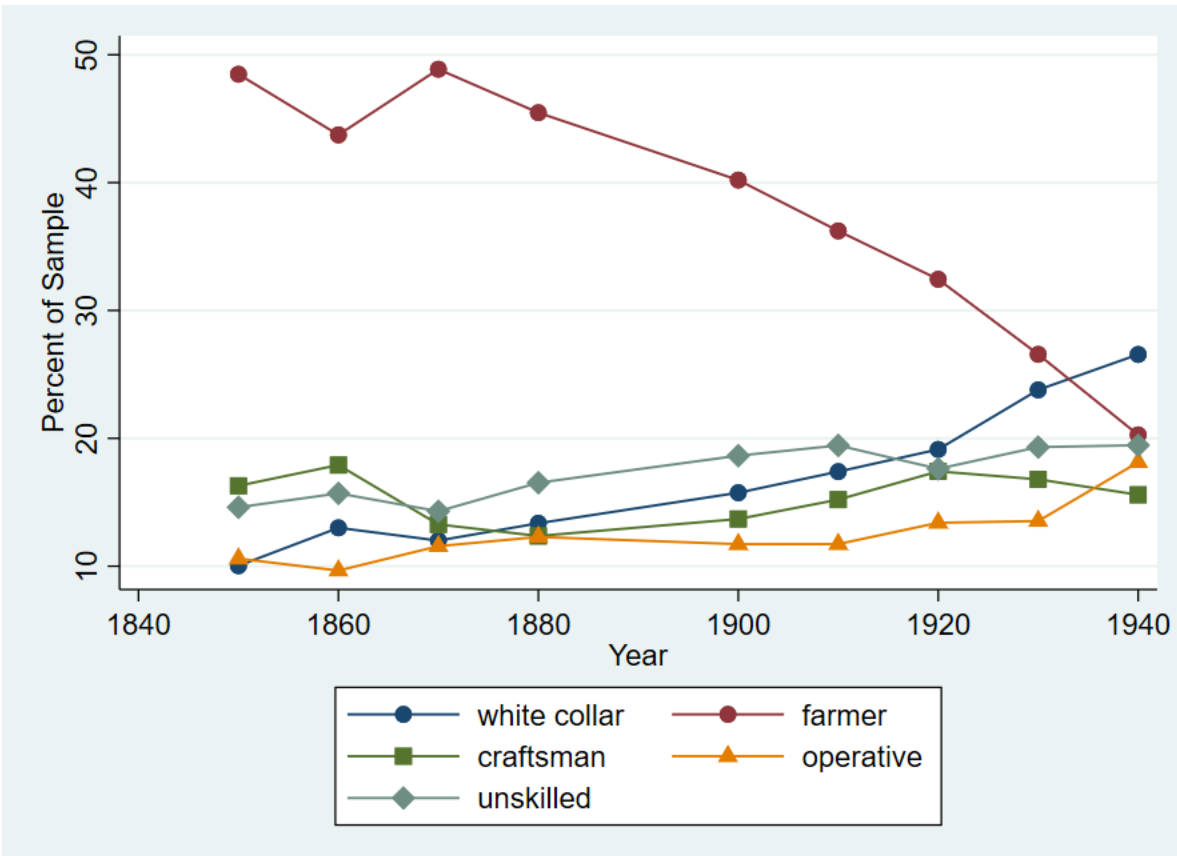

Figure 1: This figure shows the evolution of the proportion of workers in the sample over time for each of the five occupation groups. The data used are complete count US census data for every male worker with a known occupation age 16-65 living in the 48 continental United States and D.C. from IPUMS.

The primary beneficiary of this reallocation of labor away from agriculture was toward white collar work, in particular after 1920. The group with the second largest gains were from operatives. Operatives are workers who perform many of the standardized tasks associated with industrialization so it is not surprising to see their share of employment increase. In contrast, the share of workers employed as craftsmen fluctuated somewhat, but remained steady in this period. Craftsmen were the kind of workers engaged in more technical processes like cobbling that survived industrialization.

Some of these results are due to my classifying farm laborers as agricultural workers rather than as unskilled workers. With this different classification scheme the broad trends remain, but unskilled workers and white collar workers overtake farmers as the larger occupation groups by 1940. Since the ratio of farm laborers to other farm workers did not change much during this period my findings are consistent regardless of the classification one uses. This decline can even be seen in the total number of agricultural workers. While the US population increased between each year pair after 1910 the total number of agricultural workers declined in each of the following years meaning that this transition outpaced general population growth.

The same data used to generate Figure 1 can also show us how agricultural employment varied geographically. For ease of interpretation employment is shown at the census division level rather than the state level. Table 1 shows broad declines in the share of employment in agriculture in all regions except for the Pacific region if we consider 1850 our starting year. Starting in 1860 the agriculture share of each state fell by 1940.

By 1940 agricultural workers made up less than 7% of the adult male workforce in New England and the Middle Atlantic states. They made up less than 15% of the adult male workforce in the Pacific and East North Central regions. The East South Central states had the highest share of workers employed in agriculture in every year, but the two other Southern census divisions had agricultural employment shares similar to other census divisions. While the South Atlantic had over half of its workforce engaged with agriculture as late as 1910

Table 1: **Agriculture Employment Share by Census Division**

Census Division	1850	1860	1870	1880	1900	1910	1920	1930	1940
New England	36.21	31.45	27.72	22.79	16.27	13.51	10.04	8.31	6.54
Middle Atlantic	36.11	31.58	27.63	23.29	16.26	12.95	10.18	7.11	5.81
East North Central	61.34	54.08	53.48	48.24	36.94	31.55	26.06	18.69	14.66
West North Central	61.48	57.51	58.61	60.17	53.76	49.38	47.93	43.21	35.23
South Atlantic	53.61	48.12	64.54	57.44	54.05	51.01	45.08	37.20	26.83
East South Central	69.46	60.57	74.19	69.78	65.85	63.34	58.84	52.23	41.60
West South Central	47.70	50.68	66.59	64.59	65.25	60.18	54.31	46.78	35.40
Mountain	51.27	38.14	31.87	22.68	31.56	35.26	41.08	37.16	27.78
Pacific	6.21	20.76	24.93	28.16	28.76	23.45	24.33	18.19	13.64

Note: This table shows the share of the work force engaged in agriculture by census division. The data used are complete count US census data for every male worker with a known occupation age 16-65 living in the 48 continental United States and D.C. from IPUMS.

As the South Atlantic its share in agriculture fell quickly after to the point that by 1930 its share was practically identical to that of the Mountain division.

The changes in agricultural share in the West South Central division was similar to the changes seen in the East South Central, though generally higher. It was not until 1930 that the region employed less than half of its workforce in agriculture. However, rapid declines in the share after 1910 meant it caught up to the West North Central division by 1940. Since one of the key findings of this paper is that farmers face higher costs when switching into high skilled work the fact that the South had larger agricultural shares in employment during this period this may be a channel to potentially explain why the Southern labor market appeared separate from other parts of the country.

The last trend to consider using the census data for all adult male workers is the distribution of workers across space. Again, for ease of presentation these trends are shown at the census division level instead of state or county level in Table 2. Table 2 shows that even as the absolute number of workers in each division increased decade by decade there were significant changes in where workers migrated and lived. These data also reflect the increase in population in western states over time.

Table 2 shows that five census divisions lost national employment share between 1850-

Table 2: Total Work Force Share by Census Division

Census Division	1850	1860	1870	1880	1900	1910	1920	1930	1940
New England	14.47	12.06	9.97	8.46	7.16	6.83	6.77	6.29	6.32
Middle Atlantic	30.40	28.43	22.70	21.46	20.72	20.76	21.04	21.22	21.45
East North Central	22.17	24.20	23.47	22.73	21.53	20.01	20.68	20.97	20.82
West North Central	3.92	7.59	10.36	12.65	14.07	13.08	12.14	11.13	10.27
South Atlantic	13.82	11.66	14.02	13.44	12.75	12.28	12.50	12.25	12.87
East South Central	10.15	8.86	10.38	9.79	9.43	8.70	8.17	7.89	7.55
West South Central	3.14	3.97	5.04	6.29	7.80	9.16	9.53	9.98	9.54
Mountain	0.38	0.61	1.28	1.96	2.62	3.36	3.35	3.13	3.14
Pacific	1.55	2.62	2.78	3.23	3.90	5.82	5.81	7.14	8.04

Note: This table shows the share of the total work force engaged by census division. The data used are complete count US census data for every male worker with a known occupation age 16-65 living in the 48 continental United States and D.C. from IPUMS.

1940 although the declines in New England and the Middle Atlantic divisions were much greater than in the East North Central, South Atlantic, and East South Central divisions. While the Mountain division only gained about 2% of the nation's employment share during this time, its gain relative from its initial low share was greatest. The other three division all roughly gained six percentage points of national employment share.

Like all work that includes linked census data there are concerns about attrition and about the representatives of the sample [5]. If either issue is present in the linked data it will lead to improper estimation of the cost function used to measure labor market integration. In this scenario, if traits that make it easier for workers to be linked between censuses also make it easier for workers to transition between occupation groups then the cost values will underestimate the true costs. This is because workers who would have been included in the sample, but would not have moved, are being excluded.

There are no ways to avoid this problem completely, if present, and my linked data exhibit this problem. My linked samples contain anywhere from 12 to 25% of the original sample, which are standard matching rates in the literature. These samples typically over-represent farmers and under-represent unskilled workers. There is also regional biases present in each

[5] Bailey et al (2020) highlight the challenges of implementing linked data. In particular by highlighting areas were the accuracy of linked data can have serious issues.

matched pair, however the regions that are over and under represented differ across the matched pairs. The appendix provides more information detailing exactly how different the linked and original data are.

Table 3: Example Linking Table

1850-1860	**Origin**		**Destination**	
Variable	**1850 Sample**	**1850 Matched**	**1860 Sample**	**1860 Matched**
Number	5,166,780	654,722	4,384,825	654,722
White collar	10.17	10.15	13.49	13.12
Farmer	47.44	49.39	44.77	42.89
Craftsman	16.62	17.52	18.54	18.86
Operative	10.9	10.32	8.94	9.32
Unskilled	14.87	12.64	14.27	15.81
New England	14.23	13.18	12.3	10.46
Middle Atlantic	30.34	32.27	28.63	29.15
East North Central	22.26	24.00	24.15	24.89
West North Central	3.98	4.11	7.54	7.74
South Atlantic	13.77	12.71	11	11.16
East South Central	10.18	9.72	8.5	8.44
West South Central	3.22	2.66	3.93	4.14
Mountain	0.38	0.47	0.62	0.73
Pacific	1.63	0.87	2.85	3.28

Note: This table shows the results of the linking procedure used to generate the linked data. All data come from IPUMS and the Census Linking Project.

A common way to adjust for these kinds of biases is to regress matching rates on observable characteristics and generate probability weights before collapsing the individual level data to the matched cell Z_t^{ijkl} level. Using this approach generates a corrected matched sample that is more similar to the full sample. Table 3 is the linked comparison table for the 1850-1860 linked data. Only one is shown here because in all years the procedure generated similar results. For ease of interpretation, the population distribution is reported by census division rather than at the state level, but all matching was done at the state-occupation group level. The adjustment is done in a way such that neither the origin nor destination original sample data differ from the matched sample, across any of the dimensions by more than two percentage points.

The final relevant concern on the data side are the wage estimates used in equation (3).

The US census did not collect individual wage data until 1940 so two different proxy variables of wages are used in this analysis. The first proxy is known as an occupation score. These values were created by IPUMS and assign the median wage value for occupations in 1950 to every instance of the occupation in every year in my sample. The second proxy are wages interpolated from the 1940 census. Since the 1940 census includes wage data it is possible to regress individual wages on a quadratic in age-16 and use those estimated coefficients to interpolate wages in previous periods. While education and other ideal data for use in a Mincer regression are available in 1940 they are not in previous periods so only age is used. Both approaches share the flaw that they are imposing structure about wage distributions native to 1950 and 1940 back through each time period.

Each measure also introduces an additional flaw. The occupation scores assign the same value to each worker for a given occupation. Thus there is less observed wage variation within each occupation group than there actually was. The wage variation within each occupation group is now driven by the number of different occupations within it. In contrast the interpolated wages are only based on the ages of workers so they do not allow for wages to vary across locations or occupations, but do allow for variation within occupations via worker age.

Given the nature of these estimates there is a risk of attenuation bias when estimating ν_t. Since equation (1) estimates $\frac{1}{\nu_t}$ the attenuation bias would lead to ν_t being large. In particular, the attenuation bias should be worse in earlier time periods since the wage distribution in 1940 or 1950 is more likely to be similar to the distribution in 1930 than in 1850. Thus ν_t is more likely to be over-estimated in early periods. Another cause for concern is that due to data limitations this paper estimates ν_t using the simplified strategy from equation (3). Fortunately, the estimates of ν_t are similar regardless of which proxy variable for wages is used. The estimates of ν_t in this paper tend to be larger than what is traditionally found in the literature, implying there is greater variance in workers preferences for wages over amenities or that biases in the construction of the proxy variable lead workers to appear to prefer wages over amenities.

4 Results

This section presents the measures of switching costs and evaluate how they affirm the existing historical narrative around labor market integration. The results also show how the transition out of agriculture is important in understanding how labor market integration evolved in the United States during this period. First, summary statistics based on three important trends from the linked data are presented. Next the results from equation (1) are presented, followed by the results from equation (3), and lastly the point estimates of the seven relevant costs from equation (2).

Table 4: Key Statistics from Linked Data

	1850-1860	1860-1870	1870-1880	1900-1910	1910-1920	1920-1930	1930-1940
Location Changers	0.35	0.35	0.32	0.24	0.23	0.21	0.16
Job Changers	0.48	0.47	0.45	0.44	0.42	0.43	0.42
Both Changers	0.22	0.22	0.20	0.15	0.14	0.13	0.10

Note: This table shows the percent of the workers in each of the linked samples who changed location, changed job, or changed both. All data come from IPUMS and Census Linking Project.

Mobility declined broadly for workers whether migrants changed occupation, location, or both. The smallest change was for pure occupational mobility. Between 1850-1860 nearly half of the workers in the sample had changed occupation groups within ten years. By 1930-1940 the share of workers had slowly, but steadily declined to closer to 40%. The decline in the number of workers working in a different state after 10 years was more pronounced. The earliest sample period reflected the nadir of workers moving between states. By the last period the proportion of workers moving between states had declined by more than half. The same can be said for workers making both changes and it appears that the declining mobility across location drove that decline.

Based on these trends we might expect migration costs to be increasing. One might also draw the conclusion that the costs of changing location were larger than changing occupations. However, it is unwise to make a direct comparison between any of the three costs. States are large and unevenly so. This paper does not claim that the costs of changing location were larger than the costs of changing occupation. If location change patterns were the same at finer levels of geography it would be more appropriate to draw such a conclusion.

Table 5 and the rest of this section of the paper highlight the importance of including origin and destination fixed effects when studying labor market integration. The below table shows that costs declined through time, even as the share of workers making changes across locations and occupations fell. This is evidence that labor market integration was increasing even as mobility rates overall were declining. This could be because mobility within occupations or states was increasing even as mobility across states and occupations fell. It could also be because once switching costs are sufficiently low workers can make their optimal move. As costs decline workers may not move because they have already settled in their optimal cell.

Table 5: First Stage Regression Results

	1850-1860	1860-1870	1870-1880	1900-1910	1910-1920	1920-1930	1930-1940
Variables	pop	pop	pop	pop	pop	pop	pop
location_switch	-4.341***	-4.433***	-4.677***	-5.216***	-5.309***	-5.441***	-5.683***
	(0.0520)	(0.0448)	(0.0410)	(0.0339)	(0.0327)	(0.0329)	(0.0320)
occupation_switch	-1.377***	-1.467***	-1.535***	-1.542***	-1.646***	-1.628***	-1.587***
	(0.0441)	(0.0427)	(0.0412)	(0.0335)	(0.0313)	(0.0319)	(0.0324)
both_switch	0.839***	0.950***	0.962***	0.921***	1.016***	0.977***	0.873***
	(0.0587)	(0.0512)	(0.0499)	(0.0418)	(0.0397)	(0.0397)	(0.0385)
Farmer	-0.769***	-0.653***	-0.865***	-0.911***	-1.035***	-1.208***	-1.291***
	(0.0790)	(0.0734)	(0.0702)	(0.0579)	(0.0551)	(0.0548)	(0.0568)
Craftsman	-0.579***	-0.476***	-0.586***	-0.471***	-0.463***	-0.520***	-0.525***
	(0.0775)	(0.0739)	(0.0699)	(0.0611)	(0.0545)	(0.0577)	(0.0583)
Operative	-0.488***	-0.259***	-0.432***	-0.577***	-0.570***	-0.666***	-0.523***
	(0.0875)	(0.0778)	(0.0696)	(0.0653)	(0.0625)	(0.0642)	(0.0668)
Unskilled	-0.887***	-0.707***	-0.688***	-0.711***	-0.759***	-0.846***	-0.758***
	(0.0811)	(0.0740)	(0.0711)	(0.0590)	(0.0549)	(0.0606)	(0.0632)
Observations	40,700	52,320	57,360	60,025	60,025	60,025	60,025
R-squared	0.960	0.959	0.963	0.969	0.971	0.970	0.974
Origin FE	YES	YES	YES	YES	YES	YES	YES
Destination FE	YES	YES	YES	YES	YES	YES	YES

Note: This table presents results from eqn. (1) with robust standard errors included in parentheses. These include seven types of fixed costs that capture different dimensions of the choices workers make when changing occupations and locations.

Table 6 lists the estimates of $\frac{1}{\nu}$ using two proxies for wages. The first part of the table uses the occupation score variable provided by IPUMS. The second part of the table uses a wage constructed by interpolating wages using 1940 wage and age data. Both proxies share the similar weakness of imposing more rigidity than present in hypothetical true wage data. The occupation score contains rigidity within occupation description since the occupation score is the same for all workers with the same occupation description. The interpolated wages depend solely on the wage distribution by age in 1940 and enforces that structure through each of the previous time periods. With this estimate all workers with the same age earn the same interpolated wage.

Both proxy variables have limitations, but equation (3) yields similar results regardless of which is used. In both cases the estimated values or ν_t are statistically different from zero meaning that these wage estimates are picking up a trend in the fixed effects from equation (1). Regardless, the time trend of ν_t is the same for both variables. Starting with the 1870-1880 year pair ν_t declines in each year except for a small increase between 1910-1920 and 1920-1930 when using interpolated wages. These results match the original prediction for ν_t which was that it should decline over time as the greater diversity of goods leads to wages having a greater effect on utility. These values for ν_t are larger than what is commonly found in the literature. This is most likely due to differences in the availability of wage data and how the wage data were recorded.

By itself ν_t can be an interesting parameter to study, but in this paper the primary reason to estimate it is to back into the true $C_t(i, j, k, l)$ in order to understand how these costs were changing across time. Table 7 presents the fixed effects from Table 6 divided by $1\nu_t$ from Table 5. Since ν_t only varies by year pair and not by type of cost all this does is scale the costs from each column. For Table 7 values of ν_t estimated with the occupation score variable were used. The trends are the same regardless of which ν_t is used.

Table 6 lists the estimates of $\frac{1}{\nu}$ using two proxies for wages. The first part of the table uses the occupation score variable provided by IPUMS. The second part of the table uses a

Table 6: Second Stage Regression Results for Both Wage Estimates

$\frac{1}{\nu_t}$	1850-1860	1860-1870	1870-1880	1900-1910	1910-1920	1920-1930	1930-1940
Occupation Score	0.202***	0.195***	0.213***	0.258***	0.273***	0.287***	0.313***
	(0.00814)	(0.00989)	(0.00750)	(0.00646)	(0.00618)	(0.00602)	(0.00600)
Observations	185	218	239	245	245	245	245
R-squared	0.737	0.661	0.745	0.820	0.842	0.860	0.880
Interpolated Wages	0.245***	0.217***	0.247***	0.294***	0.299***	0.297***	0.309***
	(0.00651)	(0.00680)	(0.00540)	(0.00412)	(0.00397)	(0.00390)	(0.00358)
Observations	185	218	239	245	245	245	245
R-squared	0.888	0.819	0.894	0.955	0.960	0.961	0.968

Note: This table presents the results of eqn. (3) using wages interpolated from 1940 data and age characteristics from each Census and using the occupation score provided by IPUMS. The outcome variable are fixed effects from eqn. (1).

wage constructed by interpolating wages using 1940 wage and age data. Both proxies share the similar weakness of imposing more rigidity than present in hypothetical true wage data. The occupation score contains rigidity within occupation description since the occupation score is the same for all workers with the same occupation description. The interpolated wages depend solely on the wage distribution by age in 1940 and enforces that structure through each of the previous time periods. With this estimate all workers with the same age earn the same interpolated wage.

Both proxy variables have limitations, but equation (3) yields similar results regardless of which is used. In both cases the estimated values or ν_t are statistically different from zero meaning that these wage estimates are picking up a trend in the fixed effects from equation (1). Regardless, the time trend of ν_t is the same for both variables. Starting with the 1870-1880 year pair ν_t declines in each year except for a small increase between 1910-1920 and 1920-1930 when using interpolated wages. These results match the original prediction for ν_t which was that it should decline over time as the greater diversity of goods leads to wages having a greater effect on utility. These values for ν_t are larger than what is commonly found in the literature. This is most likely due to differences in the availability of wage data and how the wage data were recorded.

By itself ν_t can be an interesting parameter to study, but in this paper the primary reason to estimate it is to back into the true $C_t(i, j, k, l)$ in order to understand how these costs were changing across time. Table 7 presents the fixed effects from Table 6 divided by $1\nu_t$ from Table 5. Since ν_t only varies by year pair and not by type of cost all this does is scale the costs from each column. For Table 7 values of ν_t estimated with the occupation score variable were used. The trends are the same regardless of which ν_t is used.

Table 7: Switching Costs

Costs	1850-1860	1860-1870	1870-1880	1900-1910	1910-1920	1920-1930	1930-1940
location_switch	-21.49	-22.73	-21.96	-20.22	-19.45	-18.96	-18.16
occupation_switch	-6.82	-7.52	-7.21	-5.98	-6.03	-5.67	-5.07
both_switch	4.15	4.87	4.52	3.57	3.72	3.40	2.79
Farmer	-3.81	-3.35	-4.06	-3.53	-3.79	-4.21	-4.12
Craftsman	-2.87	-2.44	-2.75	-1.83	-1.70	-1.81	-1.68
Operative	-2.42	-1.33	-2.03	-2.24	-2.09	-2.32	-1.67
Unskilled	-4.39	-3.63	-3.23	-2.76	-2.78	-2.95	-2.42

Note: This table presents the point estimate costs from Table 3 divided by the point estimate values from Table 4 for ν_t estimated with the occupation score variable provided by IPUMS.

Unlike in Table 5 we can now directly measure how each of the seven costs evolved over time. Going row by row we first see that the cost of changing one's location declined over time and fell by 15%. The cost of changing one's occupation declined 26%. One reason for the larger decline in occupation switching costs may be due to the use of states instead of counties. The size of the rebate or cost of making both changes declined 33% meaning that workers were able to bundle costs less effectively as time went on. In other words, they kind of costs incurred by workers changing location and the costs incurred by workers changing occupation grew more different over time.

As in Table 5, the costs faced by each of the non high skilled occupation groups when transitioning into high skilled labor fell between the 1850-1860 sample and the 1930-1940 sample except for farmers. During this near century craftsman, operatives, and unskilled workers saw costs of transitioning into white collar work fall by percents of 41, 31, and 45 respectively. However, the declines were not uniform across time. For both operatives and craftsman costs increased temporarily after the Civil War before declining during the 20th century. Unskilled workers saw their costs decline regularly until 1910-1920, and then declined further during the Great Depression.

The costs for farmers instead increased by around 8% between the first and last time periods studied. The costs for farmers did not consistently decline with costs alternating between increases and decreases between each decade pair. Why these costs fluctuated may be related to the cyclical nature of farming. During the 19th and 20th centuries the US experienced many boom and bust cycles. If farming was more sensitive to these business cycle fluctuations and if a census year happened to coincide with one of those cycles it could explain this volatility.

Together the results from this paper provide evidence for increasing labor market integration. Costs were falling at the same time that labor mobility across space and occupational groups was declining. While these results may initially seem counter-intuitive, the model shows that these trends can be explained by a convergence in attractiveness of origins and destinations - in other words increasing labor market integration. The unique experience of farmers merits further investigation. Additionally, how much of Southern uniqueness with respect to national labor market integration can be attributed to its relatively high share of farm workers is an important avenue for further research.

5 Works Cited

Abramitzky, Ran, and Leah Boustan. "Immigration in American Economic History." Journal of Economic Literature, vol. 55, no. 4, 2017, pp. 1311-1345.

Artuç, Erhan, Shubham Chaudhuri, and John McLaren. "Trade Shocks and Labor Adjustment: A Structural Empirical Approach." The American Economic Review, vol. 100, no. 3, 2010, pp. 1008-1045.

Artuç, Erhan, and John McLaren. "Trade Policy and Wage Inequality: A Structural Analysis with Occupational and Sectoral Mobility." Journal of International Economics, vol. 97, no. 2, 2015, pp. 278-294.

Alston, Lee J., and T. J. Hatton. "The Earnings Gap between Agricultural and Manufacturing Laborers, 1925–1941." The Journal of Economic History, vol. 51, no. 1, 1991, pp. 83-99.

Bryan, Gharad, and Melanie Morten. "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia." Journal of Political Economy, vol. 127, no. 5, 2019, pp. 2229-2268.

Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey. "How Well Do Automated Linking Methods Perform? Lessons from US Historical Data". Journal of Economic Literature, vol 58, no 4, 2020, pp. 997-1044.

Caselli, Francesco, and Wilbur John Coleman II. "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation." The Journal of Political Economy, vol. 109, no. 3, 2001, pp. 584-616.

Collins, W., & Wanamaker, M. (2015). The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants. The Journal of Economic History, 75(4), 947-992.

Collins, Williams J. and Ariell Zimran. "Immigrants' Changing Labor Market Assimilation in the United States during the Age of Mass Migration." National Bureau of Economic Research 2019.

Correia, Sergio, Paulo Guimarães, and Thomas Zylkin. "ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects.", 2019.

Ferrie, Joseph P. "History Lessons: The End of American Exceptionalism? Mobility in the United States since 1850." The Journal of Economic Perspectives, vol. 19, no. 3, 2005, pp. 199-215.

J. M. C. Santos Silva, and Silvana Tenreyro. "The Log of Gravity." The Review of Economics and Statistics, vol. 88, no. 4, 2006, pp. 641-658.

Kim, Sukkoo. Economic Integration and Convergence: U.S. Regions, 1840-1987. The Journal of Economic History, Vol. 58, No. 3, 1998, pp. 659-683.

Mitchener, Kris J., and Ian W. McLean. "U.S. Regional Growth and Convergence, 1880-1980." The Journal of Economic History, vol. 59, no. 4, 1999, pp. 1016-1042.

Ran Abramitzky, Leah Boustan and Myera Rashid. Census Linking Project: Version 1.0 [dataset]. 2020. https://censuslinkingproject.org

Rosenbloom, Joshua L. Looking for Work, Searching for Workers: American Labor Markets during Industrialization. Cambridge University Press, Cambridge, UK;New York;, 2002.

Rosenbloom, Joshua L. "One Market Or Many? Labor Market Integration in the Late Nineteenth-Century United States." The Journal of Economic History, vol. 50, no. 1, 1990, pp. 85-107.

Salisbury, Laura. Selective Migration, Wages, and Occupational Mobility in Nineteenth Century America. Explorations in Economic History, vol. 53, 2014, pp. 40-63.

Steven Ruggles, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas and Matthew Sobek. IPUMS USA: Version 9.0 [dataset]. Minneapolis, MN: IPUMS, 2019. https://doi.org /10.18128 /D010.V9.0

Steinwender , Claudia. "Real Effects of Information Frictions: When the States and the Kingdom Became United." American Economic Review, vol. 108, no. 3, 2018, pp. 657-696.

Wright, Gavin. Old South, New South: Revolutions in the Southern Economy since the Civil War. Basic Books, New York, 1986.

Wright, Gavin. "The Economic Revolution in the American South." The Journal of Economic Perspectives, vol. 1, no. 1, 1987, pp. 161-178.

6 Appendix A: Data Construction

To construct the final samples full count census data from IPUMS for the years 1850, 1860, 1870, 1900, 1910, 1920, 1930 and 1940 are used. For each year the samples were restricted to only men who resided in the continental US. Workers living in Indian territories or in unspecified locations were excluded. All workers without a listed occupation were also excluded. The IPUMS variable *occ1950* was used to determine occupation status and all workers with a value greater than 970 were dropped. All individuals with duplicates of the IPUMS variable *histid* or who had a missing *histid* were also dropped. The occupation groups were created by assigning workers with occupation values from 1-100 and 200-499 the white collar group, workers with 100-199 and 800-899 the farmer group, workers with 500-599 the craftsman group, workers with 600-699 the operative group, and workers with values between 700-799 and 900-970 the unskilled labor group.

Next, origin and destination versions of the data for the years 1860, 1910, 1920, and 1930. It was unnecessary to do this for 1850, 1870, 1900, and 1940 because 1850 and 1900 were only ever origin years and 1870 and 1940 were only ever destination years. For all origin year samples the data were restricted to only include men age 16-55. All destination year samples restricted the data to only include men age 26-65.

Once these restrictions were in place the crosswalk for each year pair provided by the Census Linking Project (CLP) were used to link the two census years. The CLP provide links created from a variety of linking algorithms. This paper used the ABE exact standard linking procedure. Because this matched sample was not representative weights were created following the practice outlined in the readme files provided by the CLP. The only demographic variables used to generate the weights were based on five year age bins of workers, census regions, occupation groups, and worker immigrant status. Then the data were collapsed to count the number of people in each origin occupation-location and destination occupation-location pair. The data also include zero values for all empty cells.