

ORIGINAL ARTICLE

Open Access



# Immigration and wages: new evidence from the African American Great Migration

John Gardner

Correspondence:  
jgardne@olemiss.edu  
Department of Economics,  
University of Mississippi, 371  
Holman Hall, P.O. Box 1848,  
University, MS, USA

## Abstract

During the African American Great Migration, millions of blacks left the Southern USA in favor of cities in the North. Despite the social and economic consequences of this migration, the question of its impacts on labor markets in the North has largely been overlooked in the literature. In this paper, I use both local wage comparisons and structural simulations of the aggregate Northern labor market to provide new evidence on the effects of the Great Migration on wages in the North, redoubling the evidence that it caused large declines in wages for blacks, with little effect for whites. The agreement between my local and aggregate wage effect estimates has implications for our general understanding of how immigration and wages are related and how that relationship can be measured.

**JEL Classification:** J15, J61, R23, N32, N92

**Keywords:** Migration, Immigration, Internal migration, Great Migration, Local labor markets, National labor market, Wages, Spatial arbitrage

## 1 Introduction

During the African American Great Migration, millions of blacks left their places of birth in the Southern USA in favor of cities in the North. By drastically redistributing the black, this large and prolonged internal migration played an important role in shaping the economic and cultural history of the modern USA.<sup>1</sup> In particular, Smith and Welch (1989) show that the Great Migration was integral to the wage gains that blacks made relative to whites during much of the twentieth century (also see Donohue and Heckman 1991; Smith and Welch 1978). Despite its wide-reaching cultural and economic significance, the literature on the Great Migration has largely overlooked its implications for receiving cities in the North, focussing instead on the Great Migrants themselves. An exception is Boustan (2009), who finds that the Great Migration put downward pressure on the wages of blacks living in the North, with little effect on whites' wages there.

This paper redoubles the evidence on the effects of the Great Migration on wages in the North. My findings broadly support those in Boustan (2009), though they suggest that the Great Migration may have decreased the wages of Northern blacks more than previously thought. Importantly, I estimate similar effects using both a local labor markets approach that compares wages among areas with different amounts of Southern immigration and a structural, national labor market approach that combines the estimated parameters of the aggregate production function in the North with observed immigration

in order to simulate the wage effects of the Great Migration. The robustness of my findings to different methodological approaches adds credibility to the estimates and to our understanding of the consequences of the Great Migration.

In addition to shoring up the literature on the effects of the Great Migration on the North, these findings are of some significance to the broader literature on the relationship between migration (both foreign and internal) and wages. Studies of foreign immigration to the USA have not come to a strong consensus on how contemporary immigrant flows have affected natives' wages. In general, local labor markets studies tend to find smaller effects (see, for example, Altonji and Card 1991; Card 1990, 2001, 2009) while national labor market studies tend to find larger effects (see, for example, Borjas 2003, 2006; Borjas et al. 2010). The literature has identified a number of confounding factors that may explain why studies taking different methodological approaches come to different conclusions. These factors were less common during the Great Migration period, making it a useful backdrop against which to analyze the relationship between immigration and wages.

The most prominent of these explanations is the spatial arbitrage hypothesis, which holds that natives respond to immigrant inflows by migrating internally, attenuating between-labor-market supply shocks, and consequently, estimates of the wage effects of immigration that are based on geographic comparisons; empirical tests of this hypothesis have come to conflicting conclusions (cf. Borjas 2006; Card 2001; Card and DiNardo 2000; Peri and Sparber 2011). During the Great Migration period, blacks living in the North were clustered into a small number of metropolitan areas, effectively limiting their ability to move in response to inflows of Southern immigrants.<sup>2</sup> While I find evidence of a white outmigration response to Southern immigration, I find no such response among blacks living in the North. Because I also find evidence of considerable separation between the black and white labor markets in the North, this finding implies that internal migration within the North did not arbitrage away Southern-immigration-induced relative local labor supply shocks for blacks.<sup>3</sup>

Another potential explanation for the divergence between the results of local and national immigration studies is that inflows of foreign immigrants to the contemporary USA do not alter the skill distribution in a way that affects natives' wages in the long run. For example, Card (2009) argues that, since the immigrant and native skill distributions are similar for appropriately defined measures of skill, immigration has had little effect on the relative wage structure. Similarly, Ottaviano and Peri (2012) argue that imperfect substitution between US natives and foreign immigrants, which concentrates the wage impacts of immigration on immigrants themselves, can explain why structural national labor market studies that assume perfect substitution find larger wage impacts for natives. Analysis of the Great Migration does not suffer from these limitations. I provide evidence that, while within education-experience-race groups, Southern immigrants were perfectly substitutable for Northern natives during this period, blacks and whites were not. Consequently, the overwhelmingly black inflow of Southern immigrants changed the skill distribution in the North appreciably. Furthermore, over half of the black population in the North was Southern-born by 1950, implying large proportional shocks to the supply of black labor in the North with which to measure the wage effects of immigration (and by virtue of the separation between the black and white labor markets, with the potential to generate relatively large effects).

The labor market effects of the Great Migration are, like those of any immigration episode, idiosyncratic to its historical circumstances. The location decision faced by blacks born in the Southern USA, for example, differed from those faced by potential foreign immigrants to the contemporary USA; the imperfect substitution that I estimate, for another, was likely due at least in part to racial discrimination. Conclusions about the effects of the Great Migration may still be informative about those of contemporary immigration. Just as each immigration episode will impact labor markets differently, depending on the characteristics of its migrants, receiving location, macroeconomic environment, etc., the effects identified by each analysis will be specific to the episode under study. However, each new analysis helps complete our picture of the relationship between immigration and wages.

## **2 Local labor markets analysis**

### **2.1 Data and summary statistics**

The analysis in this section uses data drawn from Integrated Public Use Microdata Series (IPUMS) extracts of the 1940 and 1950 US Censuses (Ruggles et al. 2015). These data contain detailed information on wage income and residential location, as well as a host of demographic and economic variables. To my knowledge, they are the earliest data that contain rich wage and geographic information for the Great Migration period. I restrict the sample to US-born black and white men of working age and living in the North at the time of enumeration. I use the Eichenlaub et al. (2010) definition of the North and obtain similar results when the foreign born are included in the sample. I define local labor markets as Metropolitan Statistical Areas (MSAs). Further details regarding the construction of the sample and the variables used in the empirical analysis are available in Appendix 1.

Table 1 presents summary evidence on the amount of immigration from the Southern USA. Southern immigrants comprised about 8 % of the Northern labor force by 1940 and 10 % by 1950, comparable to the foreign-immigrant fraction of the modern US labor force. This immigrant flow was heavily skewed towards blacks relative to the Northern native labor force. Southern immigrants made up over 60 % of the black labor force in the North in both decades, compared to less than 10 % of the white labor force. Although the large proportion of Southern-born blacks who migrated North, and the effect of their emigration on the size of the black labor force in the North, made the Great Migration “great,” accounts of the Great Migration often overlook the fact that more Southern whites moved than blacks, with Southern whites accounting for more than 5 % of the Northern labor force, compared to only 3–4 % for blacks. The period that I study is part of the second wave of the Great Migration, which spanned roughly 1940–1970 (see Tolnay (2003) for details on the extent of Southern immigration between the full 1910–1970 period). Table 1 also gives a sense of when these immigrants arrived. In 1940, for example, less than 2 % of the Northern labor force consisted of Southern-born blacks who still resided in the South as of 1935. This suggests that the local labor markets estimates below identify a combination of long- and short-run effects of immigration.

Figure 1 plots the immigrant share of the black labor force against the share of the white labor force for Northern metropolitan areas. As the figure shows, there was substantial overlap between the destinations of black and white immigrants from the South. Table 2 compares the education distribution for natives and immigrants to the North. Southern immigrants are slightly overrepresented at lower levels of education and slightly

**Table 1** Immigration by race and year

Prop. Southern (1940)		
Entire labor force	0.079 (0.001)	0.101 (0.001)
Black labor force	0.632 (0.005)	0.661 (0.008)
White labor force	0.054 (0.001)	0.068 (0.001)
Prop. Black and Southern		
Entire labor force	0.027 (0.000)	0.036 (0.001)
Prop. White and Southern		
Entire labor force	0.052 (0.001)	0.065 (0.001)
% Recent Southern		
	1940 (5 years ago)	1950 (1 year ago)
Entire labor force	0.017 (0.000)	0.006 (0.000)
Black labor force	0.052 (0.003)	0.008 (0.001)
White labor force	0.015 (0.000)	0.006 (0.000)

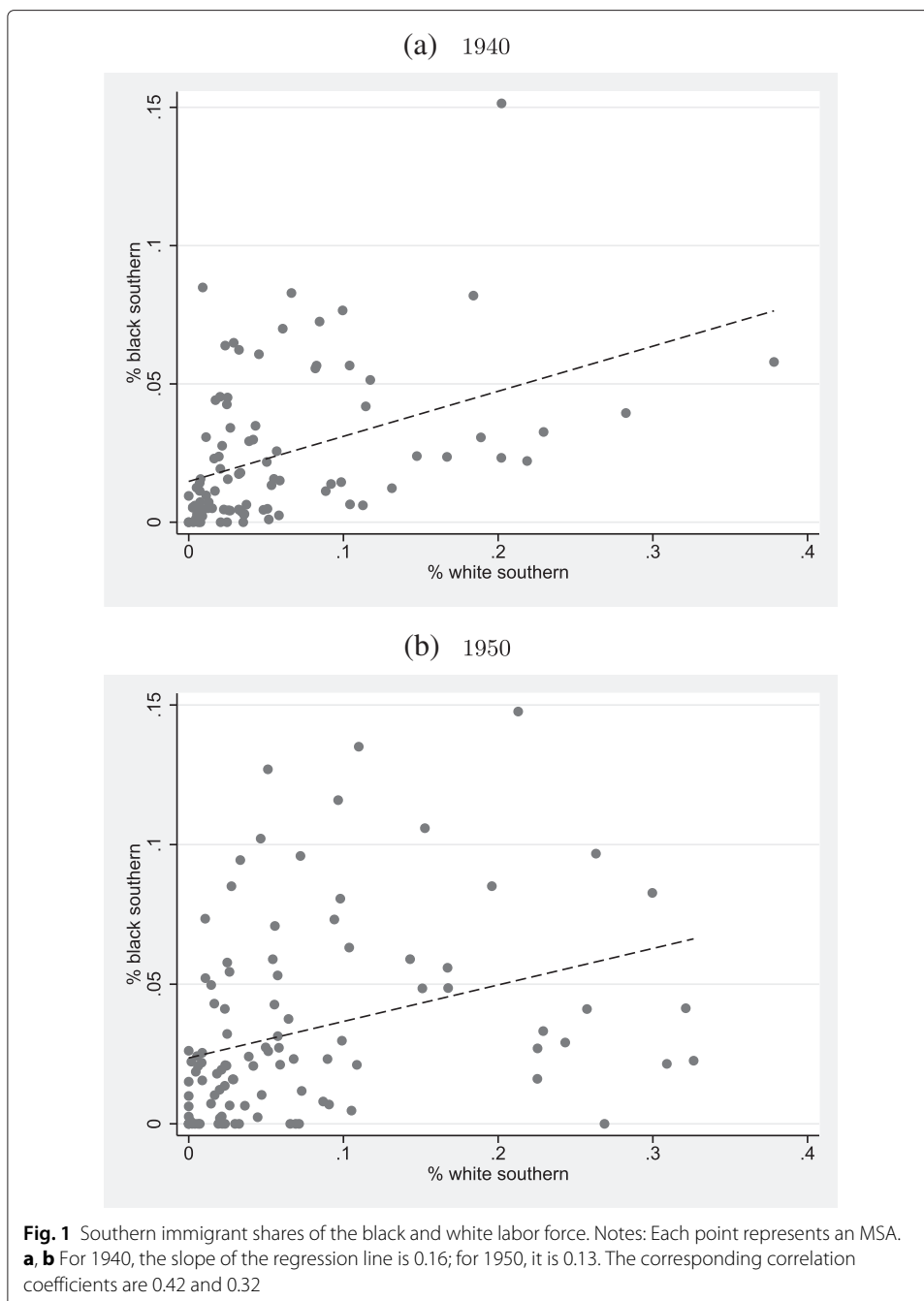
Notes: The labor force is defined as black and white men aged 16–64 who reported nonzero earnings in the year of enumeration. “Southern” means born in the South. Standard errors in parentheses

underrepresented at higher levels, particularly among blacks. However, the overall distributions are similar for both races and all years.<sup>4</sup> For this reason, I focus my analysis on race-specific average, rather than within-skill-group, effects.

## 2.2 Empirical strategy

To apply the local labor markets approach to the Great Migration, I estimate a series of linear models that relate individuals’ log wages to the immigrant share of the labor force in the metropolitan area where they live. A threat to the validity of this approach is that Southern immigrants were not randomly assigned to labor markets in the North. It is possible, for example, that immigrants selected into areas where wages were higher, perhaps due to productivity or demand shocks, in which case my estimates of the wage impacts of immigration would be attenuated. The standard solution to this problem, first used by Altonji and Card (1991), is to predict immigration according to historical immigrant settlement patterns, then use these predictions to instrument for observed immigration. This strategy draws on work by Bartel (1989), who shows that successive waves of immigrants tend to move to the same areas, a phenomenon which has also been documented among the Great Migrants (Tolnay 2003). Altonji and Card (1991) argue that this enclaving behavior provides a source of variation in immigration that is uncorrelated with the contemporary economic conditions of local labor markets.<sup>5</sup>

To exploit this variation, I predict the immigrant fraction  $p_{jt}$  of the labor force in area  $j$  and decade  $t$  using  $\hat{p}_{jt} = \mu_{j20}M_t/N_{jt}$ , where  $\mu_{j20}$  is the fraction of all Southern immigrants who resided in area  $j$  in 1920,  $M_t$  is the total number of Southerners living in the North in decade  $t$ , and  $N_{jt}$  is the total population of area  $j$  in decade  $t$ .<sup>6</sup> I then use the predicted immigrant shares  $\hat{p}_{jt}$  as instruments for the observed shares  $p_{jt}$ . In some



of the specifications that I estimate, I replace the overall immigrant fraction of the labor force with the fractions comprised of black and, separately, white Southerners, in which case the instruments are constructed by replacing the components of the  $\hat{p}_{jt}$  with their race-specific counterparts. Figure 2 plots, separately by decade and race, the immigrant share of the labor force against predictions of that share based on 1920 settlement patterns. The figure shows that immigration predicted from historical settlement patterns is strongly correlated with actual immigration. The formal first-stage regressions, presented in Appendix 3: Table 19, support this conclusion, showing that predicted immigration has

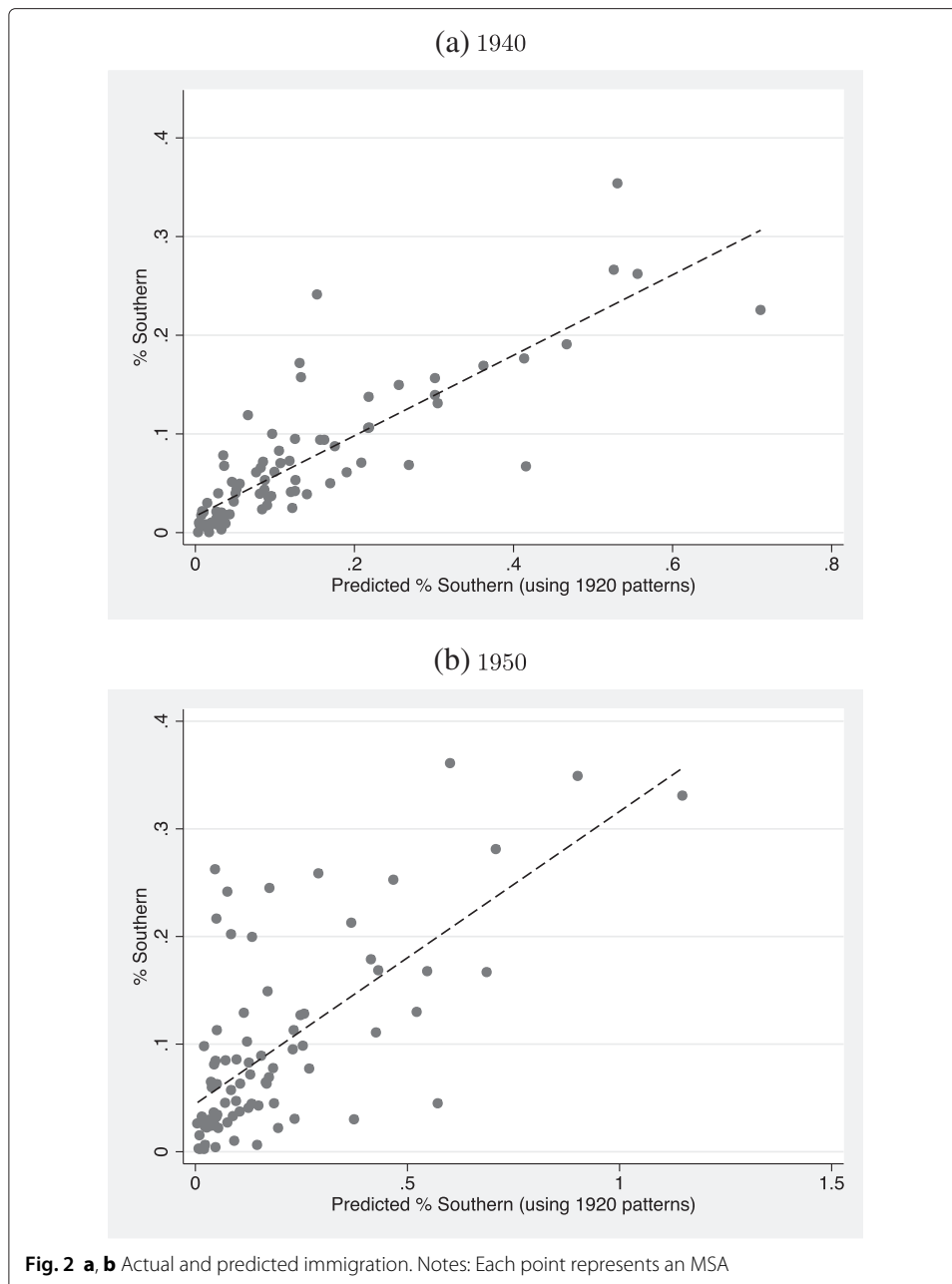
**Table 2** Education by race, year and immigrant status

Year	Education group	Black		White	
		Northern	Southern	Northern	Southern
1940	Less than 5th grade	0.10	0.16	0.02	0.05
	5th–8th grade	0.48	0.56	0.39	0.44
	Some high school	0.23	0.15	0.23	0.21
	High school degree	0.12	0.08	0.22	0.17
	Greater than high school	0.07	0.05	0.13	0.13
	Total	1.00	1.00	1.00	1.00
1950	Less than 5th grade	0.05	0.12	0.02	0.04
	5th–8th grade	0.34	0.45	0.27	0.35
	Some high school	0.28	0.24	0.24	0.24
	High school degree	0.21	0.13	0.29	0.21
	Greater than high school	0.11	0.05	0.18	0.15
	Total	1.00	1.00	1.00	1.00
1960	Less than 5th grade	0.03	0.07	0.01	0.04
	5th–8th grade	0.25	0.36	0.18	0.28
	Some high school	0.33	0.28	0.25	0.25
	High school degree	0.27	0.19	0.32	0.25
	Greater than high school	0.13	0.09	0.24	0.19
	Total	1.00	1.00	1.00	1.00

Notes: Standard errors for the estimated proportions, each of which is less than .01, suppressed

substantial power to explain the overall and race-specific immigrant shares of Northern labor markets in both the black and white regression samples.

Although the 1920 settlement patterns of Southern immigrants are necessarily uncorrelated with economic conditions that were idiosyncratic to the 1930s and 1940s in the North, instrumental variable estimates based on predicted migration patterns may still be inconsistent if metro-average wages are serially correlated and immigrants were attracted to areas that paid high wages in 1920. I take several steps to account for this possibility. To control for the local determinants of wages, I estimate specifications that include a full set of indicators for age and education, as well as the metro-level shares of the black and white populations employed in manufacturing, the shares living on farms, and metro-average education among blacks and whites. In addition, for the 1950 regression samples, I am able to include metro-average wages in 1940 as a covariate (this is not possible for the 1940 samples since the Census did not ask wage questions prior to 1940). If local wages follow a lagged dependent variable structure, including lagged average wages completely accounts for any residual correlation between wages and 1920 settlement patterns. In Appendix 1, I show that including these lags can also mitigate the inconsistency that would arise if wages follow a fixed-effects, rather than a lagged dependent variable, structure, although I also present some evidence that a lagged model fits the data better. Lagged average wages may also proxy for time-varying unobserved heterogeneity and control for any direct and lasting effect that 1920 immigration had on local wages. The models that include lagged metro-average wages should be interpreted as identifying a combination of the short- to medium-run effects of immigrant inflows and lasting effects of existing stocks of immigrants.<sup>7</sup> Because each specification models wages as a function of the overall immigrant share of the local labor force, they identify the total effect of Southern immigration on each racial group's wages.<sup>8</sup>



### 2.3 Wage impact results

Table 3 presents estimates of the impact of Southern immigration on the annual wages (all wage income earned in the year preceding enumeration) of native Northerners. The entries in the table are estimates of the coefficient on the immigrant share of the metropolitan area labor force from models that regress individual log wages on this share and other covariates. I estimate separate regressions for each racial group and Census year. For the 1940 Census samples, I estimate two specifications. Specification (1) includes individual education and age indicators as covariates; specification (2) adds the metro-level covariates described above to these individual covariates. For the 1950 samples, I also

**Table 3** Annual wages and immigration

Specification		OLS				IV			
		1940		1950		1940		1950	
		Black	White	Black	White	Black	White	Black	White
(1)	Prop. Southern	0.152 (0.207)	0.0363 (0.380)	−0.661** (0.313)	0.367 (0.270)	−0.110 (0.394)	−0.178 (0.386)	−1.116*** (0.324)	0.394 (0.276)
	Observations	2107	107,149	1043	41,154	2081	104,029	1029	39,219
	Clusters	27	93	23	100	26	80	22	80
(2)	Prop. Southern	−0.243 (0.298)	0.329 (0.261)	−0.948** (0.411)	0.475** (0.220)	−0.774*** (0.280)	0.379 (0.245)	−1.369*** (0.507)	0.367 (0.313)
	Observations	2107	104,209	1043	37,433	2081	101,339	1029	36,413
	Clusters	27	87	23	72	26	75	22	62
(3)	Prop. Southern			−1.185*** (0.343)	0.162 (0.200)			−1.827*** (0.273)	0.301 (0.251)
	Observations			1043	39,316			1029	38,078
	Clusters			23	87			22	75
(4)	Prop. Southern			−1.245*** (0.291)	0.210** (0.0993)			−2.269*** (0.546)	0.104 (0.225)
	Observations			1043	36,473			1029	35,644
	Clusters			23	67			22	59

Notes: All specifications include indicators for age and education; specification (2) includes white and black metro-level percent employed in manufacturing, percent farming, and average years of education; specification (3) includes the metro level average dependent variable, lagged one decade; specification (4) includes metro-level average variables and a lagged mean dependent variable. Regressions weighted by the number of observations used to calculate metro-level covariates. Standard errors, clustered by metro area, reported in parentheses  
 \*\*\*, \*\*, and \* denote significance at the 1, 5, and 10 % levels, respectively



estimate specifications (3) and (4), which add lagged metro-average log wages to specifications (1) and (2). The ordinary least squares (OLS) estimates are presented in the left panel of Table 3. In the 1940 sample, none of the estimated coefficients on the proportion Southern are statistically significant, although the point estimate is negative for blacks in specification (2). In the 1950 sample, the coefficient is negative and statistically significant in each specification for blacks and becomes larger in absolute value as more covariates are added to the model, a pattern consistent with selection into high-wage metropolitan areas. The coefficient on the proportion Southern is positive in all of the models estimated for whites, though it is not always statistically significant. The right panel of the table presents the instrumental variables (IV) estimates. For blacks, each IV estimate of the coefficient on the proportion Southern is larger in absolute value than its corresponding OLS estimate, also consistent with selection into high-wage areas. With the exception of specification (1) in 1940, the estimated coefficient is negative for blacks in every specification, exhibiting the same pattern of a larger absolute effect as additional covariates are added to the model. For the richest specification (specification (4), estimated using the 1950 sample), the estimate implies that a 10 % increase in the immigrant share of the labor force decreased annual wages among native blacks by about 23 %. For whites, most of the point estimates are positive and, although the estimates are more precise than those for blacks, the hypothesis that Southern immigration had no effect on wages can never be rejected.

Note that, while the wage variable used in these regressions varies at the individual level, the immigrant share varies only across metropolitan areas.<sup>9</sup> While the clustering of blacks into a small number of metropolitan areas in the North is advantageous from the standpoint of concern about spatial arbitrage, it implies that the wage effects for blacks are identified from limited variation in local immigrant shares. As a consequence, the black regression results, particularly those for the 1950 samples in which only a subset of individuals were asked wage information, should be interpreted with some caution. However, the reported standard errors are clustered at the metropolitan area, and the estimates are very similar to those obtained by regressing metro-average residual (from regressions of log annual wages on individual covariates) wages on metro-level immigrant shares and covariates (these regressions are presented in Appendix 3: Table 17).<sup>10</sup> Below, I also show that the individual-level regressions are robust to a number of tests designed to assess their sensitivity to outliers and the use of different estimation samples.

The estimates in Table 3 should be viewed as general equilibrium impacts that combine the effects of own- and cross-race immigration. In order to decompose these impacts, I replicate specifications (1)–(4), allowing the fractions of the local labor force consisting of Southern blacks and Southern whites to enter separately. Table 4 presents the estimates. Both the OLS and IV estimates suggest that the negative effect of immigration evident for blacks is driven primarily by competition between Northern-born blacks and black immigrants from the South. For native blacks' wages, the OLS estimates of the coefficients on the proportion Southern black are all negative, while those on the proportion Southern white are either positive or smaller in absolute value (none are statistically significant). The IV estimates of the coefficients on the proportion Southern black are larger in absolute value than the OLS estimates and uniformly statistically significant. The coefficients on the proportion Southern white are small and, in most cases, insignificant. The point estimate of the specification (4) coefficient implies that black immigration, which

**Table 4** Annual wages and race-specific immigration

Specification	OLS				IV				
	1940		1950		1940		1950		
	Black	White	Black	White	Black	White	Black	White	
(1)	Prop. Southern black	−0.486 (0.769)	1.943*** (0.493)	−0.702 (1.161)	1.750*** (0.279)	−1.504* (0.837)	0.708 (0.883)	−2.917** (1.245)	0.946* (0.519)
	Prop. Southern white	0.483 (0.416)	−0.579** (0.242)	−0.638 (0.401)	−0.265* (0.136)	0.730 (0.585)	−0.739** (0.301)	−0.00155 (0.606)	−0.285* (0.171)
	Observations	2107	107,149	1043	41,154	2081	104,029	1029	39,219
	Clusters	27	93	23	100	26	80	22	80
(2)	Prop. Southern black	−1.905 (1.256)	2.151*** (0.374)	−1.920 (1.482)	1.764*** (0.353)	−4.806*** (1.583)	1.734*** (0.384)	−2.265* (1.351)	0.914* (0.483)
	Prop. Southern white	0.385 (0.472)	−0.530** (0.206)	−0.535 (0.576)	−0.229 (0.189)	1.049* (0.580)	−0.557*** (0.208)	−0.646 (0.439)	−0.167 (0.268)
	Observations	2107	104,209	1043	37,433	2081	101,339	1029	36,413
	Clusters	27	87	23	72	26	75	22	62
(3)	Prop. Southern black			−1.086 (0.981)	1.330*** (0.259)			−2.659*** (0.556)	0.787** (0.325)
	Prop. Southern white			−1.241*** (0.381)	−0.262*** (0.0920)			−1.311** (0.529)	−0.272* (0.141)
	Observations			1043	39,316			1029	38,078
	Clusters			23	87			22	75
(4)	Prop. Southern black			−5.203*** (1.163)	0.845** (0.393)			−5.437*** (1.497)	0.101 (0.422)
	Prop. Southern white			−0.0541 (0.516)	−0.0360 (0.161)			−0.127 (0.487)	0.106 (0.220)
	Observations			1043	36,473			1029	35,644
	Clusters			23	67			22	59

Notes: All specifications include indicators for age and education; specification (2) includes white and black metro-level percent employed in manufacturing, percent farming, and average years of education; specification (3) includes the metro level average dependent variable, lagged one decade; specification (4) includes metro-level average variables and a lagged mean dependent variable. Regressions weighted by the number of observations used to calculate metro-level covariates. Standard errors, clustered by metro area, reported in parentheses

\*\*\*\*, \*\*\*, and \*\* denote significance at the 1, 5, and 10 % levels, respectively

increased proportion of the Northern labor force comprised of Southern blacks to 3.6 % by 1950, decreased native blacks' wages by about 15 %.

For native whites, the evidence is somewhat mixed. Both the OLS and IV estimates suggest a weak positive effect of black immigration on the wages of Northern-born whites, although this effect disappears in the IV estimate of specification (4), which includes the most covariates. At the same time, there is some evidence of a negative own-race effect for whites, with the coefficients on the white immigrant share of the labor force uniformly negative. However, these coefficients are never statistically significant when metro-level covariates are included. The estimates for whites also suggest that the apparent negative effect of immigration on the wages of native blacks cannot be explained by selection into low-wage areas: the coefficients on the black immigrant share variables are all positive, consistent with a either positive cross-race effects of immigration or selection into high-wage areas, but not selection into low-wage areas.

I also estimate models of the impact of immigration on weekly wages (annual wages divided by weeks worked). The overall and race-specific estimates, presented in Appendix 3: Tables 20 and 21, are similar to the annual wage results: Southern immigration put substantial downward pressure on native blacks' wages, an effect driven by competition from black immigrants, with no discernible impact on whites' wages. The key difference between the annual and weekly estimates is that the magnitudes of the effects of immigration on blacks' weekly wages are much smaller. In the 1950 sample, for example, the coefficient on the (overall) proportion Southern for specification (4), which includes the full set of covariates, is about  $-1.3$ , compared to about  $-2.3$  for annual wages. This difference, which implies that adjustments along the intensive margin of labor supply were an important part of the response of local labor markets to Southern immigration, helps explain the magnitude of the estimated annual wage elasticities. The prominence of adjustments along this margin also raises the possibility that both the annual and weekly wage estimates understate the true impact of Southern immigration, which may have caused some natives to exit the labor force, and consequently, my estimation samples.

These wage impact estimates suggest that black and white labor were, effectively, imperfect substitutes during the sample period. Under perfect substitution, the effects of overall, own-, and cross-race immigration would be the same for each racial group. Instead, there is a large black-white disparity in the estimated impact of overall immigration, a robustly negative effect of own-race immigration on native blacks' wages, and only tenuous, specification-sensitive evidence of an own-race effect for native whites or cross-race effects for either group. This pattern cannot be explained by outmigration; under perfect substitution, these effects would be symmetric regardless of whether outmigration attenuated relative local labor supply shocks. Nor can it be explained by measurement error in the construction of local immigrant shares, which Aydemir and Borjas (2011) show can attenuate local labor markets estimates. The estimated effect of black immigration on native blacks' wages is substantially larger than the cross-race effect even though the black samples are smaller than the white ones and therefore likely to be more error-ridden. This implies that, to the extent that my estimates are attenuated by measurement error, the underlying own-race effect for blacks must be larger than the cross-race effect.

Because Southern immigrants were disproportionately black relative to the native labor force, segmentation between the markets for black and white labor in the North would

have concentrated the effects of immigration on blacks. A simple measure of the separation between these markets is the correlation between black and white employment shares within industry-occupation cells throughout the North. This correlation, which does not reflect segregation at lower levels of aggregation, was .65 in 1940 and .41 in 1950. In the next section, I provide direct evidence of imperfect substitution: treating the entire North as an aggregate labor market, I estimate a finite elasticity of substitution between blacks and whites. The wage regressions, employment share correlations and structural elasticity estimates imply that black and white labor were utilized as though they were imperfectly substitutable, though this effective imperfect substitution may not have been due to actual productivity differences. Discrimination against blacks, for example, may have caused occupational segregation that prevented the substitution of black for white labor. Racial differences in educational attainment, and educational quality conditional on attainment, may have also rendered blacks and whites imperfect substitutes. Boustan (2009) finds that controlling for measures of educational quality can explain two thirds of the apparent imperfect substitution.

#### 2.4 Native outmigration results

As discussed in the introduction, if natives move to different labor markets in order to avoid competition from immigrants, immigration-induced labor supply shocks will be distributed across many labor markets, as will the effect of these shocks on wages. As a consequence, estimates of the impact of immigration on wages based on comparisons between local labor markets will be attenuated. Note that the possibility of spatial arbitrage among native blacks during the Great Migration is less troubling since the wage regressions reveal an appreciable effect of immigration. At worst, these regressions identify an upper bound on the underlying impact of Southern immigration on blacks' wages. On the other hand, these regressions do not suggest an own-race effect for whites or any cross-race effects. It is possible that outmigration arbitrated away local differences in these effects, even if they were present at an aggregate level in the North.

To test for outmigration, I model the propensity of natives to change locales in response to Southern immigration. Both the 1940 and 1950 Censuses provide information on individuals' recent migration behavior. The 1940 data give the metropolitan area of residence 5 years prior to enumeration while the 1950 data give the area of residence 1 year prior. For each Census decade and racial group, I regress an indicator for leaving one's pre-Census metropolitan area on the recent immigrant share of the local labor force (the number of Southerners living in the metropolitan area at the time of the Census but not in the pre-Census period as a fraction of the area's pre-Census population). Specification (1) includes age and education indicators as covariates; specification (2) includes the contemporaneous metro-average covariates used in the wage regressions and described above.<sup>11</sup>

Table 5 presents the results of this exercise. The top panel of the table contains estimates of the coefficient on the proportion Southern. The OLS estimates provide some evidence of outmigration among native whites in 1940, immigration among native blacks in that decade, and little evidence of outmigration among natives of either race in 1950. The IV estimates, which use the same instrument as the wage regressions, show a clear pattern of outmigration for whites in 1940 but no evidence of black outmigration in any decade or statistically significant evidence of white outmigration in 1950. The bottom panel of

**Table 5** Native outmigration and recent immigration

		OLS				IV			
		1940		1950		1940		1950	
		Black	White	Black	White	Black	White	Black	White
Overall									
Specification (1)									
	Prop. Southern	-0.428 (0.351)	0.343*** (0.102)	-0.784 (1.149)	0.423 (0.315)	-0.0468 (0.267)	1.101*** (0.346)	-0.219 (0.277)	0.513 (0.757)
	Observations	2018	97,043	954	38,073	1996	94,370	940	36,456
	Clusters	27	92	21	97	26	79	20	79
Specification (2)									
	Prop. Southern	-0.580*** (0.187)	-0.0283 (0.0223)	0.411 (1.091)	-0.304 (0.313)	0.223 (0.609)	0.725*** (0.215)	0.295 (2.028)	1.356 (0.876)
	Observations	2018	94,512	954	34,942	1996	92,054	940	34,055
	Clusters	27	86	21	70	26	74	20	61
By race									
Specification (1)									
	Prop. Southern black	0.250 (1.119)	-1.165 (0.956)	-2.655** (1.248)	-0.679 (0.958)	1.506 (3.607)	2.440 (2.420)	1.854 (5.019)	-2.992 (3.348)
	Prop. Southern white	-0.472 (0.404)	0.728*** (0.206)	0.826 (0.592)	1.051*** (0.369)	-0.593 (1.011)	0.794*** (0.295)	-1.167 (1.993)	2.340*** (0.853)
	Observations	2018	97,043	954	38,073	1996	94,370	940	36,456
	Clusters	27	92	21	97	26	79	20	79
Specification (2)									
	Prop. Southern black	0.349 (2.669)	-0.589 (0.586)	-2.735 (2.155)	-0.290 (1.290)	7.386 (6.773)	0.558 (1.532)	2.627 (4.570)	-0.157 (3.247)
	Prop. Southern white	-0.603* (0.307)	0.492*** (0.152)	-0.319 (1.287)	0.527 (0.438)	-1.697 (1.128)	0.769** (0.361)	-1.193 (1.562)	2.125*** (0.723)
	Observations	2018	94,512	954	34,942	1996	92,054	940	34,055
	Clusters	27	86	21	70	26	74	20	61

Notes: All specifications include indicators for age and education; specification (2) includes white and black metro-level percent employed in manufacturing, percent farming, and average years of education. Regressions weighted by the number of observations used to calculate metro-level covariates. Standard errors, clustered by metro area, reported in parentheses  
 \*\*\*\*, \*\*\*, and \*\* denote significance at the 1, 5, and 10 % levels, respectively

the table contains estimates of the coefficients on race-specific immigrant shares. Consistent with imperfect substitution between black and white labor, both the OLS and IV estimates show that white outmigration was a response to Southern white immigration. In particular, in the white samples, each IV estimate of the coefficient on the proportion Southern white is positive and statistically significant, while none of the coefficients on the proportion Southern black are. None of the estimates evince outmigration by native blacks.

Since some of the outmigration that took place from 1935 to 1940 and 1949 to 1950 was likely in response to competition from immigrants who arrived before these periods began, the estimates in Table 5 may overstate native whites' outmigration response.<sup>12</sup> This effect may be less pronounced in the 1940 samples since a longer period of observation is more likely to capture natives' response to recent immigration. For example, the specification (2) IV estimate of the white outmigration response to white immigration is 2.13 in the 1950 sample, while the 1940 estimate of .77 is more reasonable, though it still implies a considerable outmigration response.

In summary, there is no evidence of black outmigration, but strong evidence of a native white response to Southern white immigration. My local labor markets estimates of the impact of Southern immigration on native blacks' wages are therefore unattenuated by spatial arbitrage. Although my estimates do not suggest that outmigration among native whites completely offset incoming Southern whites, the possibility remains that spatial arbitrage obscures an own-race effect for whites or cross-race effects for both groups (e.g., if inflows of whites combined with black-white imperfect substitution put upward pressure on blacks' wages). Since immigration only increased the supply of white labor in the North modestly, the underlying wage impacts may have been small and difficult to detect even absent outmigration, making the influence of spatial arbitrage on my estimates for whites unclear. However, the national labor market analysis below results in similarly small cross-race and white own-race effect estimates, suggesting that spatial arbitrage does not explain why the white wage regression coefficients are small.<sup>13</sup>

## 2.5 Robustness and sensitivity tests

The instrumental variable estimates presented so far have been based on the total fraction of each metropolitan area's labor force comprised of Southern immigrants in 1920. That is, they are based on the propensity of Southerners to migrate to areas where other Southerners have gone. Although predicted immigrant shares based on these settlement patterns are strongly related to actual shares, a more compelling behavioral story might be that immigrants originating from a particular locale prefer to migrate to areas where others from that locale have moved. I also estimate models that instrument for immigration using shares predicted from state-specific historical settlement patterns. I predict the decade- $t$  fraction of the labor force in area  $j$  comprised of Southern immigrants using  $\hat{p}_{jt} = \sum_k \mu_{kj20} M_{kt} / N_{jt}$ , where  $\mu_{kj20}$  is the fraction of immigrants from state  $k$  that resided in metropolitan area  $j$  in 1920,  $M_{kt}$  is the number of immigrants originating from  $k$  in  $t$ , and  $N_{jt}$  is the population of area  $j$  in decade  $t$ . Using these state-specific migration patterns does not substantively change the results. The annual wage effect estimates, presented in Table 6, show that immigration decreased blacks' wages, that this decrease was a consequence of competition from Southern black immigrants, and that whites' wages were not affected by the arrival of Southern immigrants. Although the state-specific IV estimates

**Table 6** Annual wages and immigration: state-specific IV estimates

Specification	Overall				By race					
	1940		1950		1940		1950			
	Black	White	Black	White	Black	White	Black	White		
(1)	Prop. Southern	−0.308 (0.494)	−0.448 (0.394)	−1.134*** (0.307)	0.269 (0.288)	Prop. Southern black	−2.212* (1.160)	0.584 (0.834)	−2.443** (1.112)	1.155** (0.503)
	Observations	2081	104,029	1029	39,219	Prop. Southern white	0.694 (0.765)	−0.841*** (0.291)	−0.347 (0.545)	−0.303** (0.149)
	Clusters	26	80	22	80	Observations	2081	104,029	1029	39,219
						Clusters	26	80	22	80
(2)	Prop. Southern	−0.804*** (0.307)	0.121 (0.259)	−1.348*** (0.365)	0.325 (0.308)	Prop. Southern black	−4.391*** (1.325)	1.518*** (0.419)	−2.803* (1.522)	1.091** (0.460)
	Observations	2081	101,339	1029	36,413	Prop. Southern white	0.809* (0.486)	−0.644*** (0.219)	−0.640 (0.444)	−0.195 (0.276)
	Clusters	26	75	22	62	Observations	2081	101,339	1029	36,413
						Clusters	26	75	22	62
(3)	Prop. Southern			−1.570*** (0.233)	0.260 (0.198)	Prop. Southern black			−2.057*** (0.702)	1.067*** (0.309)
	Observations			1029	38,078	Prop. Southern white			−1.300*** (0.446)	−0.240* (0.125)
	Clusters			22	75	Observations			1029	38,078
						Clusters			22	75
(4)	Prop. Southern			−1.650*** (0.331)	0.244 (0.198)	Prop. Southern black			−4.907*** (1.335)	0.370 (0.461)
	Observations			1029	35,644	Prop. Southern white			0.0116 (0.471)	0.160 (0.195)
	Clusters			22	59	Observations			1029	35,644
						Clusters			22	59

Notes: All specifications include indicators for age and education; specification (2) includes white and black metro-level percent employed in manufacturing, percent farming, and average years of education; specification (3) includes the metro level average dependent variable, lagged one decade; specification (4) includes metro-level average variables and a lagged mean dependent variable. Regressions weighted by the number of observations used to calculate metro-level covariates. Standard errors, clustered by metro area, reported in parentheses  
 \*\*\*\*, \*\*\*, and \*\* denote significance at the 1, 5, and 10 % levels, respectively

for blacks are larger in absolute value than the OLS estimates in Table 3, they are also smaller than the IV estimates in that table, which are based on the 1920 settlement patterns of all Southern immigrants. One explanation for this difference is that, if it is less costly to migrate to closer areas, state-specific settlement patterns may partially identify areas to which residents of a particular state are likely to move in response to transient wage shocks. If so, historical settlement patterns for all Southerners will be less contaminated with selection effects since they better-identify migrations motivated by enclaving behavior.<sup>14</sup>

Although 1920 settlement patterns are necessarily uncorrelated with the conditions of local labor markets in the North that were idiosyncratic to 1940 and 1950, if 1920 immigrants systematically selected into Northern labor markets with above- or below-average wages and the factors that determine local average wages are serially correlated, immigrant shares predicted from these settlement patterns may be spuriously correlated with wages in future periods. In the wage regressions above, I account for this possibility by estimating models that include lagged metro-average wages. The evidence presented so far is consistent with selection into high wage areas: including covariates or instrumenting for immigration decreases the estimated wage effects, and the coefficients on the proportion Southern black in the native white regression samples are positive. These patterns also suggest that any remaining selection leads my wage regressions to understate the effects of immigration, in which case they can be interpreted as upper bounds on the underlying effects.

Since it is not clear a priori whether a lagged dependent variable structure or a fixed-effects one provides a better description of the process that determines area-level wages over time, I also pool the 1940 and 1950 data in order to estimate models that include metropolitan area fixed effects. The estimates are presented in the top panel of Table 7. Although including fixed effects reduces their precision, the estimates are consistent with the previous wage regressions: immigration decreased wages for native blacks but had no discernible effect on whites' wages (note that the hypothesis of a positive effect can be rejected for three of the four black IV regressions). The bottom panel of the table contains estimates of regressions of decadal changes in metro-level mean residual wages (i.e., after removing the variation explained by individual age and education) on decadal changes in immigrant shares. The first-differenced models preserve the variation in the area-level covariates explained by the individual-level covariates, increasing the precision of the estimates, which are broadly similar to the fixed-effects estimates.

Both the fixed-effects and first-differenced estimates suggest much larger wage effects for blacks than the decade-specific wage regressions discussed previously, although given their standard errors they are also consistent with the main regression results presented above. One explanation for this difference is that, because these models use the variation in changes in immigrant shares (rather than levels, as the decade-specific models use), they identify the effects of recent immigration, which may be larger if labor markets adapt to immigration over time, for example through endogenous adjustments to capital stock or industrial composition. However, the magnitudes of the point estimates may also indicate that the fixed-effects and first-differenced models are misspecified. As I show in Appendix 1, if wages follow a lagged dependent variable structure, estimates of a misspecified model that assumes a first-differenced/fixed-effects structure will be biased down. The intuition for this result is that if the initial wave of immigrants is attracted to high



**Table 7** Fixed-effects and first-differenced models

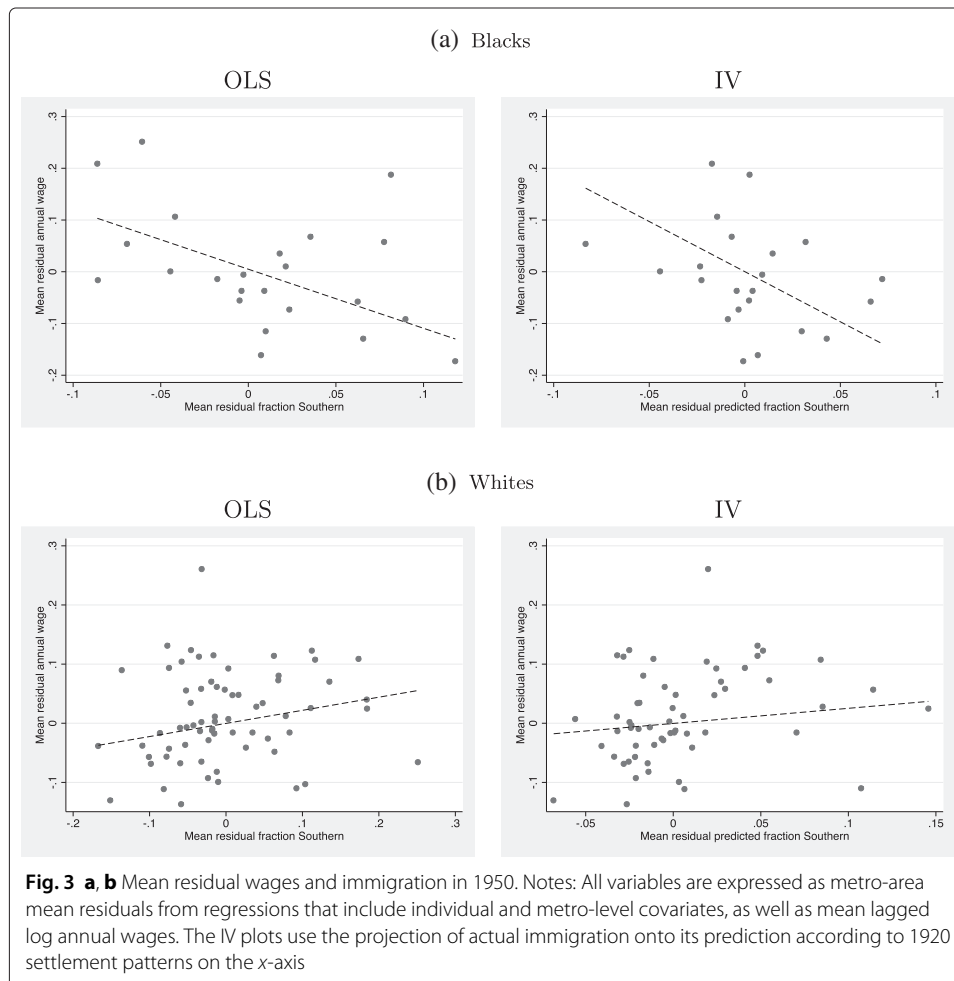
		OLS				IV			
		Black		White		Black		White	
Fixed effects									
Annual wages	Prop. Southern	-0.276 (1.948)	0.212 (1.345)	1.354*** (0.405)	1.413*** (0.299)	-6.675* (3.731)	-3.420 (5.360)	-1.373 (2.993)	-0.862 (4.749)
	Observations	3150	3150	148,303	141,642	3110	3110	143,248	137,752
	Clusters	29	29	100	92	28	28	80	78
	Covariates	N	Y	N	Y	N	Y	N	Y
Weekly wages	Prop. Southern	-1.193 (0.839)	-0.488 (1.053)	0.847** (0.366)	0.908*** (0.309)	-4.339 (2.817)	-6.411 (4.691)	-0.217 (1.680)	0.958 (4.202)
	Observations	3135	3135	147,855	141,232	3095	3095	142,825	(0.839)
	Clusters	29	29	100	92	28	28	80	78
	Covariates	N	Y	N	Y	N	Y	N	Y
First differences									
Annual wages	Prop. Southern	-0.188 (2.154)	0.445 (2.150)	0.254 (0.253)	0.297 (0.372)	-9.796*** (2.584)	-7.385** (2.905)	17.00 (135.7)	-6.143 (8.574)
	Observations	21	21	93	67	20	20	59	59
	Covariates	Y	N	Y	N	Y	N	Y	N
Weekly wages	Prop. Southern	-0.144 (1.261)	0.154 (1.294)	0.0978 (0.190)	0.217 (0.275)	-5.490** (2.215)	-5.104** (2.262)	15.35 (122.0)	-6.362 (8.788)
	Observations	21	21	93	67	20	20	59	59
	Covariates	Y	N	Y	N	Y	N	Y	N

Notes: Covariates include white and black metro-level percent employed in manufacturing, percent farming and average years of education and well as indicators for age and educational attainment. The dependent variable for the first-differenced models is the mean residual from a regression on indicators for age and educational attainment. Standard errors for the fixed-effects estimates are clustered at the metro level, for the first-differenced estimates they are heteroskedasticity robust

\*\*\*\*, \*\*\*, and \*\* denote significance at the 1, 5, and 10 % levels, respectively

wage areas, future waves are attracted to areas with established enclaves, and wages are mean reverting, areas with declining wages will receive more immigrants, inducing a spurious correlation between *changes* in wages and *changes* in immigration. This possibility makes the lagged dependent variable estimates, which can be interpreted conservatively as upper bounds, preferable.

During the period that I study, blacks were clustered into a small number of metropolitan areas in the North. For example, since the 1950 Census only asked wage questions of a subset of households, my sample for that year is relatively small, containing only 22 metropolitan areas with at least ten native blacks meeting the other selection criteria. This clustering is advantageous from an identification standpoint since it eliminates concerns about spatial arbitrage among blacks working in the North. However, since it also implies that the effects of Southern immigration on blacks' wages are identified from variation across a limited number of labor markets, it also raises the concern that the estimated effects are artifacts of a small number of outlying metropolitan areas. Figure 3, which plots metro-average residual wages against residual immigration for blacks and whites in 1950, presents graphical evidence that this is not the case.<sup>15</sup> These plots show that neither group's estimated effect is driven by outliers.



A related concern is that the racial difference in the estimated effects of immigration is a consequence of the dispersion of whites across more labor markets. For example, if whites were more likely than blacks to live in areas where the effect of immigration was small, the racial difference in the estimated wage effects could be due to sample selection. To test for this, I replicate the white (annual) wage regressions, restricting the samples to the same areas used to estimate the black wage regressions. The results, presented in Table 8, show that this restriction does not eliminate the racial gap in wage impact estimates (note that the results for blacks in this table are identical to the annual wage results for blacks given in Tables 3 and 4). Table 9, which replicates the outmigration results using the same sample restriction, shows that the estimated racial difference in outmigration behavior cannot be explained by geographical heterogeneity, either.

As a further test of the influence of the clustering of the black population on my estimates, I estimate wage and outmigration regressions using samples that include Southern immigrants. Even if interest centers on the impacts of immigration on the entire Northern labor force, there is an argument for excluding immigrants from the estimation samples. There may be labor-market-specific factors that affect immigrants' wages (e.g., if some areas are more receptive to immigrants) and the enclaving behavior that justifies my instrumental variable strategy may arise because immigrants fare better in areas where previous waves of immigrants have settled (e.g., if previous immigrants help new ones find jobs; see White and Lindstrom (2005)). In either case, including immigrants in the estimation sample may introduce endogeneity between immigration and wages. With this caveat in mind, the main advantage of including immigrants in the samples is that, since immigrants comprised over half of the Northern black labor force, doing so increases the number of metropolitan areas that can be used to estimate the effects of immigration. Table 10 presents the IV regression results for annual wages. Including immigrants increases the number of metropolitan areas used for blacks from 26 to 47 in 1940 and from 22 to 33 in 1950. The magnitudes of the estimated effects are smaller than when only natives are included in the sample, although they point to the same broad conclusion as the native regressions—black immigration decreased blacks' wages and neither white nor black immigration affected whites' wages. The outmigration results, given in Table 11, are nearly identical to those obtained using only natives.

I also estimate the annual wage regressions using two additional samples in order to further assess the sensitivity of my results to the number of metropolitan areas included in the regression samples (in the interest of brevity, I do not present these results, though they are available upon request). The first sample includes both US- and foreign-born men. Including the foreign born slightly increases the number of metropolitan areas in the black regression samples (to 30 in 1940 and 23 in 1950), and the overall pattern of the estimates is similar. The point estimates on the proportion Southern are slightly smaller, possibly because foreign-born blacks are imperfect substitute for US-born blacks, and therefore less affected by competition from Southern immigration. The second sample relaxes the exclusion of metropolitan areas with fewer than ten native born blacks or whites. When this restriction is relaxed, the black wage regressions include observations from 87 metropolitan areas in 1940 and 72 areas in 1950. Because the additional metropolitan areas contain few native blacks, the results are nearly identical to those presented in Table 3.

**Table 8** Annual wages and immigration, same metro areas

Specification	Overall				By race					
	1940		1950		1940		1950			
	Black	White	Black	White	Black	White	Black	White		
(1)	Prop. Southern	−0.110 (0.394)	−0.422 (0.424)	−1.116*** (0.324)	0.239 (0.302)	Prop. Southern black	−1.504* (0.837)	−0.132 (1.040)	−2.917** (1.245)	0.687 (0.608)
	Observations	2081	73,638	1029	26,562	Prop. Southern white	0.730 (0.585)	−0.598* (0.354)	−0.00155 (0.606)	−0.295 (0.195)
	Clusters	26	26	22	22	Observations	2081	73,638	1029	26,562
						Clusters	26	26	22	22
(2)	Prop. Southern	−0.774*** (0.280)	0.252 (0.240)	−1.369*** (0.507)	0.164 (0.385)	Prop. Southern black	−4.806*** (1.583)	0.923 (0.582)	−2.265* (1.351)	0.522 (0.616)
	Observations	2081	73,638	1029	26,562	Prop. Southern white	1.049* (0.580)	−0.0756 (0.187)	−0.646 (0.439)	−0.0986 (0.265)
	Clusters	26	26	22	22	Observations	2081	73,638	1029	26,562
						Clusters	26	26	22	22
(3)	Prop. Southern			−1.827*** (0.273)	0.316 (0.318)	Prop. Southern black			−2.659*** (0.556)	0.963*** (0.324)
	Observations			1029	26,562	Prop. Southern white			−1.311** (0.529)	−0.357*** (0.131)
	Clusters			22	22	Observations			1029	26,562
						Clusters			22	22
(4)	Prop. Southern			−2.269*** (0.546)	−0.0549 (0.241)	Prop. Southern black			−5.437*** (1.497)	−0.398 (0.678)
	Observations			1029	26,562	Prop. Southern white			−0.127 (0.487)	0.171 (0.184)
	Clusters			22	22	Observations			1029	26,562
						Clusters			22	22

Notes: Instrumental variables estimates. All specifications include indicators for age and education; specification (2) includes white and black metro-level percent employed in manufacturing, percent farming, and average years of education; specification (3) includes the metro level average dependent variable, lagged one decade; specification (4) includes metro-level average variables and a lagged mean dependent variable. Regressions weighted by the number of observations used to calculate metro-level covariates. Standard errors, clustered by metro area, reported in parentheses

\*\*\*\*, \*\*\*, and \*\* denote significance at the 1, 5, and 10 % levels, respectively

**Table 9** Native outmigration and recent immigration, same metro areas

Specification	Overall				By race					
	1940		1950		1940		1950			
	Black	White	Black	White	Black	White	Black	White		
(1)	Prop. Southern	-0.0468 (0.267)	1.214*** (0.394)	-0.219 (0.277)	0.572 (0.804)	Prop. Southern black	1.506 (3.607)	4.393 (3.363)	1.854 (5.019)	-2.366 (3.320)
	Observations	1996	68,564	940	24,704	Prop. Southern white	-0.593 (1.011)	0.493* (0.299)	-1.167 (1.993)	2.173*** (0.824)
	Clusters	26	26	20	20	Observations	1996	68,564	940	24,704
						Clusters	26	26	20	20
(2)	Prop. Southern	0.223 (0.609)	0.768*** (0.226)	0.295 (2.028)	1.355 (1.013)	Prop. Southern black	7.386 (6.773)	6.803 (5.551)	2.627 (4.570)	-0.254 (3.497)
	Observations	1996	68,564	940	24,704	Prop. Southern white	-1.697 (1.128)	-0.597 (1.038)	-1.193 (1.562)	1.995*** (0.653)
	Clusters	26	26	20	20	Observations	1996	68,564	940	24,704
						Clusters	26	26	20	20

Notes: Instrumental variables estimates. All specifications include indicators for age and education; specification (2) includes white and black metro-level percent employed in manufacturing, percent farming, and average years of education. Regressions weighted by the number of observations used to calculate metro-level covariates. Standard errors, clustered by metro area, reported in parentheses  
 \*\*\*\*\*, \*\*\*, and \* denote significance at the 1, 5, and 10 % levels, respectively

**Table 10** Annual wages and immigration, including Southern workers

Specification	Overall				By race					
	1940		1950		1940		1950			
	Black	White	Black	White	Black	White	Black	White		
(1)	Prop. Southern	−0.0649 (0.319)	−0.172 (0.364)	−0.567* (0.306)	0.302 (0.257)	Prop. Southern black	−0.653 (0.879)	0.788 (0.809)	−1.046 (0.963)	0.846* (0.511)
	Observations	6849	109,298	3501	41,794	Prop. Southern white	0.274 (0.490)	−0.751*** (0.286)	−0.235 (0.447)	−0.314* (0.162)
	Clusters	47	80	33	80	Observations	6849	109,298	3501	41,794
					Clusters	47	80	33	80	
(2)	Prop. Southern	−0.441* (0.228)	0.361 (0.245)	−0.469* (0.285)	0.293 (0.307)	Prop. Southern black	−3.777*** (1.076)	1.780*** (0.370)	−0.652 (0.806)	0.788 (0.511)
	Observations	6849	106,507	3501	38,906	Prop. Southern white	0.895*** (0.338)	−0.612*** (0.195)	−0.326 (0.352)	−0.165 (0.272)
	Clusters	47	75	33	62	Observations	6849	106,507	3501	38,906
					Clusters	47	75	33	62	
(3)	Prop. Southern			−0.880* (0.466)	0.183 (0.268)	Prop. Southern black			−0.891 (0.645)	0.649* (0.353)
	Observations			3488	40,602	Prop. Southern white			−0.874** (0.442)	−0.315** (0.151)
	Clusters			32	75	Observations			3488	40,602
					Clusters			32	75	
(4)	Prop. Southern			−0.620** (0.278)	0.0207 (0.213)	Prop. Southern black			−1.731** (0.673)	−0.0630 (0.445)
	Observations			3488	38,104	Prop. Southern white			0.0790 (0.299)	0.0982 (0.224)
	Clusters			32	59	Observations			3488	38,104
					Clusters			32	59	

Notes: Instrumental variables estimates. All specifications include indicators for age and education; specification (2) includes white and black metro-level percent employed in manufacturing, percent farming, and average years of education; specification (3) includes the metro level average dependent variable, lagged one decade; specification (4) includes metro-level average variables and a lagged mean dependent variable. Regressions weighted by the number of observations used to calculate metro-level covariates. Standard errors, clustered by metro area, reported in parentheses  
 \*\*\*\*, \*\*\*, and \*\* denote significance at the 1, 5, and 10 % levels, respectively

**Table 11** Outmigration and immigration, including Southern workers

Specification		Overall				By race				
		1940		1950		1940		1950		
		Black	White	Black	White	Black	White	Black	White	
(1)	Prop. Southern	0.109 (0.198)	1.027*** (0.341)	-0.210 (0.383)	0.719 (0.860)	Prop. Southern black	0.205 (2.083)	2.146 (2.236)	-2.917 (4.304)	-1.985 (3.736)
	Observations	6335	98,378	3286	38,680	Prop. Southern white	0.0777 (0.519)	0.763*** (0.269)	1.051 (1.566)	2.097** (0.880)
	Clusters	46	79	31	79	Observations	6335	98,378	3286	38,680
						Clusters	46	79	31	79
(2)	Prop. Southern	-0.0623 (0.317)	0.680*** (0.226)	-0.676 (1.444)	1.807 (1.247)	Prop. Southern black	-0.939 (6.423)	0.418 (1.518)	-0.949 (3.239)	1.258 (4.524)
	Observations	6335	95,987	3286	36,217	Prop. Southern white	0.160 (1.341)	0.750** (0.355)	-0.461 (1.088)	2.088** (0.919)
	Clusters	46	74	31	61	Observations	6335	95,987	3286	36,217
						Clusters	46	74	31	61

Notes: Instrumental variables estimates. All specifications include indicators for age and education; specification (2) includes white and black metro-level percent employed in manufacturing, percent farming, and average years of education. Regressions weighted by the number of observations used to calculate metro-level covariates. Standard errors, clustered by metro area, reported in parentheses  
 \*\*\*\*, \*\*\*, and \*\* denote significance at the 1, 5, and 10% levels, respectively

### 3 National labor market analysis

In this section, I apply a national labor market approach to the Great Migration, treating the North as one aggregate labor market. An advantage of the national approach is that it is less susceptible to the possibility of spatial arbitrage since attrition out of aggregate labor markets is less common than internal migration within them. Although my local labor markets evidence implies that blacks in the North did not respond to Southern immigration through outmigration, a comparison of local and national estimates for blacks provides a unique opportunity to cross validate these approaches in a setting where the empirical evidence against outmigration is supported by historical context. The wage regressions discussed above do suggest white outmigration, and comparing the local and national wage estimates for whites can clarify the extent to which they are attenuated by spatial arbitrage. Though the purpose of my implementation of this approach is to compare national and local estimates of the effect of Southern immigration on wages during comparable time periods, it can also be viewed in part as a replication of Boustan (2009).

My implementation follows Borjas (2003), Ottaviano and Peri (2012) and, in the Great Migration context, Boustan (2009). Because this methodology is well-developed in the literature, I sketch the implementation in the text and present additional details in Appendix 2. I assume that aggregate output in the North is produced using capital and labor according to a Cobb-Douglas technology. The labor input  $L$  consists of education-group-specific labor supplies  $L_e$ , aggregated using a constant elasticity of substitution structure with productivity weights  $\theta_e$  and elasticity  $\sigma_e$ :

$$L = \left( \sum_e \theta_e L_e^{\frac{\sigma_e-1}{\sigma_e}} \right)^{\frac{\sigma_e}{\sigma_e-1}}.$$

Analogously, the  $L_e$  consist of the labor  $L_{ex}$  supplied by those with the same education but different amounts of labor market experience, aggregated according to a CES structure with weights  $\theta_{ex}$  and elasticity of substitution  $\sigma_{ex}$ . The  $L_{ex}$  consist of the labor  $L_{exr}$ ,  $r \in \{b, w\}$ , supplied by blacks and whites with the same education and experience, aggregated using weights  $\theta_{exr}$  and elasticity  $\sigma_r$ . Finally, the  $L_{exr}$  are CES-aggregated from education-experience-race-birthplace supplies  $L_{exri}$ ,  $i \in \{s, n\}$ , using weights  $\theta_{exri}$  and elasticity  $\sigma_i$ .

Under this nested-CES structure, the effects of immigration on wages can be expressed in terms of skill-group-specific labor supplies, the wage-bill shares accruing to different groups, and group-specific elasticities. Although they are not directly observable, under the assumption that they are time-invariant, the elasticities can be estimated from time series on group-specific wages and labor supply (in principle, the productivity weights can vary over time).

#### 3.1 Data

My national labor market results are also based on extracts of the US Census provided by IPUMS (Ruggles et al. 2015). To make the national and local estimates comparable, I use the same sample restrictions for each approach (US-born black and white men living in the North, aged 16–64, with nonzero wages in the year before enumeration). I add the 1960 Census to the estimation sample in order to generate time-series variation in wages and labor supply, which the national approach requires to estimate the elasticities.<sup>16</sup> In order to compare my results to those in Boustan (2009), who studies the



1940–1970 period, I also use samples that include the 1970 Census. However, my focus is on comparing my national and local estimates, for which the shorter sample is more appropriate.

Table 2 shows that average education was low during the period that I study, making the schemes used by studies of contemporary immigration to assign individuals to education groups inappropriate. Instead, I divide the sample into five education groups: less than fifth grade, fifth-to-eighth grade, some high school, high school, and greater than high school. I impute labor market experience as age minus six minus the number of years of schooling an individual in the midpoint of the education group would have received and assign individuals to one of eight 5-year experience groups. See Appendix 2 for further detail.

The most disaggregated labor nest is the education-experience-race-birthplace group. In this nest, I measure labor supply as the number of individuals in the group and wages as group-average annual wages (note that this allows me to capture adjustments to the intensive margin of labor supply). Labor supplies and wages for higher nests are calculated as implied by the nested-CES structure (see Appendix 2).

### 3.2 Elasticity estimates

Under the nested-CES production function and competitive pricing, immigrant-native relative wages satisfy

$$\log\left(\frac{w_{exrnt}}{w_{exrst}}\right) = \log\left(\frac{\theta_{exrnt}}{\theta_{exrst}}\right) - \frac{1}{\sigma_i} \log\left(\frac{L_{exrnt}}{L_{exrst}}\right). \quad (1)$$

I use this relationship to recover the native-immigrant elasticity of substitution,  $\sigma_i$ , by fitting models of the form

$$\log\left(\frac{\bar{w}_{exrit}}{\bar{w}_{exrst}}\right) = \lambda_{exr} + \lambda_{ert} + \lambda_{xrt} + \frac{1}{\sigma_i} \log\left(\frac{\hat{L}_{exrnt}}{\hat{L}_{exrst}}\right) + u_{exrt}, \quad (2)$$

where  $\bar{w}_{exrit}$  is the average wage paid to members of an education-experience-race-birthplace group,  $\hat{L}_{exrit}$  is the (calculated) size of the group,  $\lambda_{exr}$  is an education-experience-race-group fixed effect,  $\lambda_{ert}$  is a decade-specific education-race-group fixed effect, and  $\lambda_{xrt}$  is a decade-specific experience-race-group fixed effect. The fixed effects are included to absorb as much variation as possible in the relative productivity terms. Note that in these models, only within-education-experience-race group variation over time in relative labor supply can be endogenous—this would require that, for example, only in some year, black immigrants with a high school degree and 5 years of experience supplied more labor than their native counterparts, or more immigrant than native blacks with 5 years of experience completed high school, in response to a productivity shock that favored immigrants, but black immigrants with different skills and otherwise-similar white immigrants did not.<sup>17</sup>

Table 12 presents the estimates. To allow for the possibility that the immigrant-native elasticity varies by race, I estimate separate models for blacks and whites as well as a model that constrains the elasticity to be the same over a pooled sample of both races. Regardless of the period examined or the racial composition of the sample, none of the coefficient estimates are significantly different from zero, providing no indication of imperfect substitution between Southern- and Northern-born workers. These are the only parameters

**Table 12** Elasticity estimates (immigrant-native)

Sample		1940–1960	1940–1970
Whites	$-\sigma_i^{-1}$	−0.05 (0.06)	−0.05 (0.04)
	$N$	117	156
Blacks	$-\sigma_i^{-1}$	0.05 (0.08)	0.02 (0.05)
	$N$	117	156
Pooled	$-\sigma_i^{-1}$	−0.04 (0.06)	−0.04 (0.04)
	$N$	234	311

Notes: Except where noted, wages and labor supplies are measured using all Northern labor. Standard errors for the estimates of  $\sigma_i$ ,  $\sigma_r$ , and  $\sigma_x$  are clustered by education-experience group (or race-education-experience groups for models that pool both races); standard errors for  $\sigma_e$  are heteroskedasticity-robust. All regressions are weighted by the number of observations used to construct the dependent variable

in my national labor market analysis for which I do not reject the null of perfect substitution, despite being based on the largest samples and having some of the most precise coefficient estimates.

A similar theoretical argument shows that the elasticity of substitution,  $\sigma_r$ , between black and white labor can be recovered from estimates of

$$\log\left(\frac{\bar{w}_{exwt}}{\bar{w}_{exbt}}\right) = \lambda_{ex} + \lambda_{et} + \lambda_{xt} - \frac{1}{\sigma_r} \log\left(\frac{\hat{L}_{exwt}}{\hat{L}_{exbt}}\right) + u_{ext}, \tag{3}$$

where  $\bar{w}_{exrt}$  and  $\hat{L}_{exrt}$  are education-experience-race group wages and labor supplies. Based on the evidence that immigrants and natives are perfect substitutes, I construct the  $\hat{L}_{exrt}$  as the sum of immigrant and native labor supplies within education-experience groups in each decade.<sup>18</sup> To account for potentially endogenous black-white relative labor supply within education-experience groups, I use two instrumental variable strategies, both based on the premise that Southern immigration shifts relative supply curves. The first, following Borjas (2003) and Ottaviano and Peri (2012), is the immigrant component of black-white relative labor supply within education-experience groups. The second, following Boustan (2009), is the ratio of blacks to whites among the national stock of Southerners within education-experience groups (i.e., among all Southerners, regardless of whether they have migrated North). The immigrant component instrument is appropriate if within-group labor supply is endogenous with respect to productivity shocks that affect the North and South alike, since these shocks would not have attracted immigrants. The national stock instrument is more appropriate for productivity shocks local to the North, which would be unlikely to elicit a labor-supply response among many Southerners. Although my immigrant-native elasticity estimates imply that skill-group-average wages should be calculated using all workers, I also estimate (3) using only natives' wage to allow for the possibility that immigrant-biased productivity shocks affected black-white relative labor supply.

The results, given in Table 13, point to considerable imperfect substitution between blacks and whites in both sample periods. In the 1940–1960 sample, the estimated coefficients on relative labor supply cluster around  $-.2$ , or an elasticity of 5. Only when the equation is estimated using the entire wage sample and the immigrant component instrument is the coefficient estimate indistinguishable from zero, applying the same

**Table 13** Elasticity estimates (black-white)

Wage sample		1940–1960			1940–1970			
		OLS	IV (Immig. component)	IV (National stock)	OLS	IV (Immig. component)	IV (National stock)	
All	$-\sigma_r^{-1}$	−0.22 (0.09)	−0.10 (0.11)	−0.20 (0.10)	$-\sigma_r^{-1}$	−0.13 (0.06)	−0.05 (0.08)	−0.16 (0.06)
	<i>N</i>	117	117	117	<i>N</i>	155	155	155
Natives	$-\sigma_r^{-1}$	−0.18 (0.09)	−0.19 (0.11)	−0.16 (0.13)	$-\sigma_r^{-1}$	−0.14 (0.05)	−0.12 (0.07)	−0.16 (0.07)
	<i>N</i>	117	117	117	<i>N</i>	156	155	156

Notes: "Immig. component" refers to the immigrant component of labor supply and "National stock" refers to the labor supply among all Southern-born workers (including those living in the North). Except where noted, wages and labor supplies are measured using all Northern labor. Standard errors for the estimates of  $\sigma_i$ ,  $\sigma_r$ , and  $\sigma_x$  are clustered by education-experience group (or race-education-experience groups for models that pool both races); standard errors for  $\sigma_e$  are heteroskedasticity-robust. All regressions are weighted by the number of observations used to construct the dependent variable

instrument to the native wage sample yields a significant estimate of  $-.19$ . When 1970 is added to the sample, the coefficient estimates are smaller in absolute value (between  $-.13$  and  $-.16$ ), indicating a larger elasticity of substitution between black and white labor during the 1960s. As in the shorter sample, for all but one of the estimates can perfect substitution be rejected.<sup>19</sup>

Under a nested-CES structure, shocks to white education-experience-group labor supplies that hold labor supplies in higher nest constant only affect whites' wages if blacks and whites are imperfect substitutes, implying that the black-white elasticity can be estimated independently from samples of only whites or only blacks. To provide further evidence on this elasticity, I estimate versions of (2) in racial-group-specific levels rather than black-white ratios, reporting the results in Table 14. For both sample periods, the coefficient estimates for whites (the coefficients on log labor supply) are larger in absolute value than those obtained when the equation is estimated in black-white ratios, while the implied coefficients for blacks (obtained by adding the coefficients on log labor supply and  $\text{black} \times \log \text{ labor supply}$ ) are considerably smaller. This pattern, which is consistent with greater endogeneity of black than white labor supply, explains why the coefficient estimates are smaller when the equation is estimated in ratios: they represent a weighted average of the black and white coefficients. Note that even the estimates obtained by applying the immigrant component instrument to all whites' wages are larger in absolute value and significantly different from zero. As when the equation is estimated in ratios, adding 1970 to the sample shrinks the estimated coefficients, implying greater black-white substitutability. However, the effect of adding 1970 is smaller in Table 14. For example, the coefficient estimated using OLS and all wages increases from  $-.25$  to  $-.23$ , a change more consistent with racial progress during the 1960s than the more severe changes evident in Table 12. This difference may arise because, in 1970, black labor supply remained relatively endogenous but the black labor force was larger and distributed over a greater number of education-experience cells, exacerbating the attenuation caused by estimating the equation in ratios.

As I note in the introduction, although occupational segregation arising from discrimination may have contributed to effective imperfect substitution between blacks and whites, analysis of the Great Migration can still provide externally valid insight into the effects of an influx of immigrants with a different skill mix than the native population. Discrimination may also create the appearance of imperfect substitution if, as in Becker (1971), heterogeneity in employers' tastes for discrimination meant that blacks were paid less in labor markets or skill groups containing more blacks. In this case, the Great Migration would only be informative about contemporary flows of immigrants who belong to racial, ethnic, or other minorities. The race-specific elasticity estimates in Table 14 are inconsistent with this hypothesis. If the estimated elasticities were an artifact of heterogeneous tastes for discrimination, we would expect them to be driven primarily by variation in black labor supply and wages; we observe the opposite.

To estimate the elasticity of substitution,  $\sigma_x$ , between education-experience groups, I use the (conservative) OLS estimates of (3) to compute the education-experience-group  $\times$  year labor supplies and average wages implied by the nested CES model. I then regress (logs of) these wages on education-time and experience-time fixed effects and log labor supply by OLS and by IV using the immigrant component of labor supply. I use the resulting IV estimates to compute the implied education-group  $\times$  year labor supplies

**Table 14** Race-specific estimates of  $\sigma_r$ 

Wage sample	Labor supply sample		1940–1960			1940–1970		
			OLS	IV (Immig. component)	IV (National stock)	OLS	IV (Immig. component)	IV (National stock)
All	All	log(L)	−0.25 (0.06)	−0.17 (0.05)	−0.26 (0.05)	−0.23 (0.04)	−0.17 (0.03)	−0.22 (0.04)
		Black*log(L)	0.17 (0.08)	0.07 (0.06)	0.25 (0.08)	0.15 (0.05)	0.09 (0.04)	0.18 (0.05)
		N	234	234	234	311	311	311
Native	All	log(L)	−0.26 (0.06)	−0.18 (0.05)	−0.26 (0.05)	−0.23 (0.04)	−0.18 (0.03)	−0.22 (0.04)
		Black*log(L)	0.19 (0.09)	0.06 (0.08)	0.27 (0.12)	0.12 (0.06)	0.06 (0.05)	0.14 (0.06)
		N	234	234	234	312	311	312

Notes: The estimating equation is  $\log w_{ext} = \lambda_{exr} + \lambda_{ert} + \lambda_{ext} - \sigma_w^1 \log L_{ext} - (\sigma_b^{-1} - \sigma_w^{-1}) 1_{r=b} \log L_{ext} + u_{ext}$ . Immig. component refers to the immigrant component of labor supply and National stock refers to the labor supply among all Southern-born workers (including those living in the North). Standard errors for the estimates of  $\sigma_i$ ,  $\sigma_r$ , and  $\sigma_x$  are clustered by education-experience group (or race-education-experience groups for models that pool both races). All regressions are weighted by the number of observations used to construct the dependent variable

and average wages in order to recover the elasticity of substitution,  $\sigma_e$ , between education groups from regressions of log wages on time fixed effects, education-group-specific linear time trends and log labor supply, both by OLS and by IV using the immigrant component instrument aggregated to the education-group level (Appendix 2 provides the details and theoretical motivation for these procedures). The estimates from these models are presented in Table 15. My estimates of  $\sigma_x^{-1}$  lie in the interval [.18, .26], comparable to those in Welch (1979), Card and Lemieux (2001), Borjas (2003), Boustan (2009), and Ottaviano and Peri (2012). There is more variation in estimates of  $\sigma_e^{-1}$  across the literature; mine are in the interval [.21, .28], comparable to those in Boustan (2009) and Ottaviano and Peri (2012), but larger than those in Borjas (2003) (time period differences may help explain why I find different education groups more easily substitutable).

### 3.3 Wage impact results

Under the nested-CES production function, when immigrants and natives are perfect substitutes and capital has fully adjusted to immigrant-induced labor supply shocks, the effect of immigration into all education-experience-race groups ( $e', x', r'$ ) on the competitive wage paid to labor belong to group ( $e, x, r$ ) can be expressed as

$$\begin{aligned} \frac{\Delta w_{exr}}{w_{exr}} = & \frac{1}{\sigma_e} \sum_{e'} \sum_{x'} \sum_{r'} s_{e'x'r's} \frac{\Delta L_{e'x'r's}}{L_{e'x'r's}} + \left( \frac{1}{\sigma_x} - \frac{1}{\sigma_e} \right) \sum_{x'} \sum_{r'} \frac{s_{ex'r's}}{s_e} \frac{\Delta L_{ex'r's}}{L_{ex'r's}} \\ & + \left( \frac{1}{\sigma_r} - \frac{1}{\sigma_x} \right) \sum_{r'} \frac{s_{exr's}}{s_{ex}} \frac{\Delta L_{exr's}}{L_{exr's}} - \frac{1}{\sigma_r} \frac{s_{exrs}}{s_{exr}} \frac{\Delta L_{exrs}}{L_{exrs}}, \end{aligned} \tag{4}$$

where  $\Delta L_{exrs}/L_{exrs}$  is the proportional change in immigration to group ( $e, x, r$ ),  $s_{exr}$  is the share of the wage bill accruing to skill group ( $e, x, r$ ),  $s_{ex}$  is the share of the wage bill accruing to the aggregate group with shared skills ( $e, x$ ), and so on.<sup>20</sup> Expression (4) clarifies the channels through which immigration affects wages in a nested-CES economy: as long as substitutability increases with nest depth, immigrants depress wages for those who share their skills and increase wages for those who do not.<sup>21</sup>

I use (4) to compute the effects of Southern immigration from 1940–1960 and 1940–1970 on wages in the North. Skill-group-specific proportional changes in immigration

**Table 15** Elasticity estimates (education and experience)

		1940–1960		1940–1970	
		OLS	IV (Immig. component)	OLS	IV (Immig. component)
Experience	$-\sigma_x^{-1}$	-0.26	-0.20	$-\sigma_x^{-1}$	-0.23
		(0.03)	(0.03)		(0.04)
	<i>N</i>	117	117	<i>N</i>	155
Education	$-\sigma_e^{-1}$	-0.24	-0.28	$-\sigma_e^{-1}$	-0.21
		(0.14)	(0.06)		(0.10)
	<i>N</i>	15	15	<i>N</i>	20

Notes: "Immig. component" refers to the immigrant component of labor supply and "National stock" refers to the labor supply among all Southern-born workers (including those living in the North). Except where noted, wages and labor supplies are measured using all Northern labor. Standard errors for the estimates of  $\sigma_i$ ,  $\sigma_r$ , and  $\sigma_x$  are clustered by education-experience group (or race-education-experience groups for models that pool both races); standard errors for  $\sigma_e$  are heteroskedasticity-robust. All regressions are weighted by the number of observations used to construct the dependent variable

are directly observable (I list them in Appendix 3: Table 22). To assess the sensitivity of wage-effect simulations to the racial elasticity of substitution, I use the values of  $\sigma_r$  implied by both the OLS and national-stock IV estimates of (4) obtained using the full wage sample. I use the IV estimates of  $\sigma_x$  and  $\sigma_e$ . While it was typical for a Northern worker to have less than a high school education in 1940 (see Table 2), the proportional changes in immigration were larger in more-educated skill groups, reflecting increasing educational attainment over time. Using 1940 wage-bill shares to compute (4) may therefore understate the effects of immigration over the 20- and 30-year periods that I analyze by underweighting skill groups that were initially small. A simple solution is to weight skill groups by their wage-bill shares in 1950, the midpoint of the shorter sample period, though I also report results obtained using 1940 shares. I weight skill-group-specific effects by wage-bill shares in order to present education-race, race, and overall average wage impacts (this weighting scheme is suggested by Ottaviano and Peri (2012)).

I summarize the wage impact simulations in Table 16. When skill groups are weighted by their 1940 wage-bill shares, the calculations imply that immigration between 1940 and 1960 decreased the average wage paid to black labor by about 14 % and had no effect on the average white wage; the calculations are insensitive to whether the OLS or IV estimates of  $\sigma_r$  are assumed. Using 1950 wage-bill shares, which put more weight on skill groups with larger proportional changes in immigration, gives a decline for blacks of about 20 % on average and a negligible increase for whites.<sup>22</sup> Over the longer period spanning 1940–1970, increases in substitutability of black for white labor offset increases in black immigration, resulting in average wage impacts similar to those for the shorter period, regardless of the weighting scheme. When I use 1940 wage-bill shares as weights, the average wage impact calculations are close to those reported by Boustan (2009).

The average wage decline for blacks was driven by even larger declines among blacks with at least some high school. Less-educated black labor aggregates, which received

**Table 16** The effects of South-North migration on wages

		1940–1960				1940–1970			
		1940 shares		1950 shares		1940 shares		1950 shares	
		OLS	IV	OLS	IV	OLS	IV	OLS	IV
White	Overall	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.01
Black	Overall	−0.14	−0.13	−0.22	−0.20	−0.09	−0.11	−0.19	−0.24
White	Less than 5th grade	0.00	0.00	0.02	0.02	0.03	0.03	0.05	0.05
	5th–8th grade	0.02	0.02	0.03	0.03	0.03	0.03	0.05	0.05
	Some high school	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
	High school degree	0.00	0.00	0.00	0.00	−0.01	−0.01	−0.01	−0.01
	Greater than high school	−0.01	−0.01	0.01	0.01	−0.03	−0.03	−0.01	−0.01
Black	Less than 5th grade	0.03	0.03	0.04	0.04	0.09	0.10	0.12	0.13
	5th–8th grade	−0.03	−0.03	−0.02	−0.02	0.05	0.05	0.06	0.06
	Some high school	−0.33	−0.30	−0.40	−0.36	−0.22	−0.27	−0.30	−0.37
	High school degree	−0.43	−0.39	−0.53	−0.49	−0.46	−0.57	−0.67	−0.82
	Greater than high school	−0.31	−0.28	−0.28	−0.25	−0.39	−0.48	−0.36	−0.44

Notes: Columns labeled “OLS” use the OLS estimate of the elasticity of substitution between blacks and whites; columns labeled “IV” use estimates of this elasticity obtained by instrumenting for relative labor supplies with relative supplies among the national stock of Southerners. All other elasticities are taken from IV estimates based on the labor supplied within skill groups by immigrants. “1940 shares” means that immigration and wage effects are weighted by 1940 wage bill shares, and similarly for “1950 shares”

few Southern immigrants, saw trivial changes in wages. Because immigrants comprised a small fraction of the white labor force, the proportional changes in white immigration were relatively small, and the white immigrant and native skill distributions were broadly similar, the effects of Southern immigration were small across all white labor aggregates. For the same reasons, the Great Migration would have had an equally negligible impact on blacks' wages had black and white labor been perfectly substitutable.

#### **4 Conclusions**

Comparing local labor markets, I find that Southern immigration decreased the wages paid to blacks working in the North substantially, an effect which was driven by competition from Southern black immigrants. At the same time, I find no evidence that immigrants of either race had any effect on whites' wages in the North. These findings are robust to several alternative, non-causal interpretations. Nor do I find evidence of an outmigration response among native blacks to Southern immigration. Although I do find evidence of outmigration among whites, my estimates do not imply that outmigration offset immigration entirely, suggesting that spatial arbitrage alone cannot explain the lack of an apparent wage effect for whites.

My national labor market estimates are highly consistent with these findings. By 1950, about 10 % of the Northern labor force consisted of Southern immigrants. The estimates of my richest local labor markets regression specification implies that a 10 % increase in the immigrant share of the labor force caused a 20 % decline in native blacks' annual wages. In my preferred national labor market simulation, immigration to the North between 1940 and 1960 (or 1970) lead to a 20 % decline in blacks' wages.<sup>23</sup> Like the local labor markets estimates, the national market simulations reveal that the effects of Southern immigration on whites' wages were negligible. These small simulated effects provide additional evidence that spatial arbitrage does not explain the small wage regression coefficients for whites. They also emphasize the importance of imperfect substitution between blacks and whites: had these groups been perfectly substitutable, the Great Migration would have had little effect on the distribution of skill, and hence the wage structure, in the North.

The magnitudes of the wage impacts that I estimate for blacks, and the mutual concordance between my local and national estimates and my findings with Boustan's, are unusual compared to studies of foreign immigration to the contemporary USA. These differences arise because of several factors that set the Great Migration apart from other immigration episodes, making it a potentially informative counterfactual to contemporary immigration. My wage-regression and structural estimates both suggest effective imperfect substitution between blacks and whites, and the summary evidence and elasticity estimates that I present support this conclusion. Because Southern immigrants were disproportionately black relative to the native population of the North, this imperfect substitution meant that the Great Migration altered the distribution of skill in the North appreciably. In addition, since Southern immigrants and Northern natives were perfect substitutes, these changes were not concentrated on existing immigrants. In contrast, contemporary foreign immigrant and US native skill distributions are much more similar (Card 2009), as were the within-racial-group distributions of Southern immigrant and Northern native skill during the Great Migration. If more Southern whites had moved north, or if those that did had different skills than white natives, Southern immigration



would have had a larger impact on Northern whites' wages. These counterfactual impacts, however, are not well-identified from the actual variation in white immigration and wages, demonstrating the need to examine multiple migration episodes in order to develop a complete picture of the relationship between immigration and wages.

My wage effect estimates for blacks demonstrate that immigration can, and did, alter the wage structure substantially—that relative demand curves are downward sloping, even in the long run. They also demonstrate that, when immigration does affect wages, its effects can be identified by local labor markets comparisons and structural simulations of the national economy. On the other hand, my estimates for whites show that some immigrant flows do not affect wages—that, while relative demand is downward sloping, immigration does not necessarily increase relative labor supply curves.

### Endnotes

<sup>1</sup> Tolnay (2003) provides an excellent overview.

<sup>2</sup> Although identifying the underlying cause of this racial difference in outmigration behavior is beyond my scope, historical context suggests a natural explanation: that the same social forces that contributed to segregation within cities (see Massey and Denton 1993) also operated at larger geographic scales during the Great Migration period. As part of my instrumental variables strategy, I show that Southern-born blacks shared the well-known tendency of immigrants to move to areas where previous immigrants have settled (see Bartel 1989), providing some evidence that preferences for the racial composition of residential areas influenced blacks' location decisions.

<sup>3</sup> Boustan et al. (2010), using a different identification strategy, also provide evidence of a mobility response to immigration during this period (their pooled dataset consists mostly of whites).

<sup>4</sup> Although the 1960 IPUMS does not identify MSAs and cannot be included in my local labor markets analysis, I also include education statistics for 1960 since I do use data for that year in the national labor market analysis below.

<sup>5</sup> The instrumental variables estimates therefore identify an average wage effect of inflows of immigrants who would not have migrated absent existing enclaves, insofar as these immigrants impact labor markets differently than others.

<sup>6</sup> I use 1920 settlement patterns for several reasons. First, the Great Migration did not begin until roughly 1915, so prior to 1920 the number of Southern immigrants residing in the North is small. Second, metropolitan areas are only identified in the IPUMS Census samples if they are sufficiently populous to meet confidentiality requirements, so using earlier settlement patterns reduces the number of areas that can be matched to lagged migration. Third, as Table 1 shows, many of the Southern immigrants living in the North in 1940 were already present by 1935; immigrants living in the North in 1940 may therefore have selected into locations on the basis of local economic conditions during the 1930s, so predicted migration based on 1930 settlement patterns may be more contaminated with selection effects.

<sup>7</sup> Because the number of Southern immigrants in the North was still small in 1920, this is not much of a concern (see Tolnay et al. 2005). The period that I study (1940–1970) is part of the larger second wave of the Great Migration. Nevertheless, including lagged wages restores the validity of the instrument even if the small waves of early immigrants exerted a strong impact on Northern labor markets.

<sup>8</sup>This is in contrast to models of skill-group-specific wages as a function of skill-group-specific immigration, which only identify what Ottaviano and Peri (2012) call the “direct partial effect” of immigration (Dustmann et al. 2016, make a similar point).

<sup>9</sup>I use individual-level regressions for efficiency (Lang and Gottschalk 1996) and to ensure that the standard errors appropriately reflect the removal of variation in the dependent variables explained by the included individual-level covariates (Dumont et al. 2005).

<sup>10</sup>The standard errors for the metro-level estimates are comparable to, and in many cases smaller than, the clustered standard errors for the microdata regressions.

<sup>11</sup>Peri and Sparber (2011), investigating why different tests for spatial arbitrage in the contemporary US lead to different conclusions, use simulated data to assess the performance of different specifications of internal migration behavior. My specification is a variation on their preferred one that uses microdata and models gross, rather than net, outmigration.

<sup>12</sup>A back-of-the-envelope calculation confirms this intuition. Replacing the recent immigrant share with the total immigrant share and estimating specification (2) using the 1950 data produces a coefficient of about 0.06 for native whites. Since recent immigrants comprised roughly 15 % of the total immigrant population in 1950, if this coefficient only reflected recent immigration, the implied coefficient on the recent immigrant share would be 0.4 ( $= .06/.15$ ). The actual coefficient is about 1.4, implying that only roughly 30 % of the 1949–1950 native outmigration occurred as a response to immigrants who arrived during this period.

Also note that if the coefficients are interpreted as the *within period* response to immigration, they definitely overstate the outmigration response since some of the observed outmigration must be a response to pre-period immigration if outmigration occurs with a lag. If they are interpreted as the *total* response to immigration, they will overstate the response if immigration is slowing (which it was during sample period) since there will be more observed outmigration in response to previous immigration than unobserved (due to truncation) outmigration in response to recent immigration. An alternative specification would include both the stock and flow of immigrants—such a specification is challenging to estimate since (i) enclaving behavior induces a strong correlation between the stock and the flow and (ii) IV estimation requires separate instruments for the stock and the flow.

<sup>13</sup>A natural question is why the white outmigration response was large if the wage effects were small. One possibility is that there were short-run effects which are not well-identified by either the local or national approach. Another is that white immigration had a greater effect on employment than wages. Another still is that white natives moved in response to perceived effects which never materialized.

<sup>14</sup>Another possibility is that the state-specific patterns are measured with more error, attenuating the IV estimates towards the OLS ones.

<sup>15</sup>The residuals are obtained from projections onto individual and area covariates, including lagged average wages. For the IV plot, the projection of actual onto predicted immigration is used in place of immigration itself.

<sup>16</sup>I do not use this period for the local labor markets analysis because MSA-level data are not publicly available for 1960.

<sup>17</sup>Also note that the estimated average impact of immigration on wages in higher nests will not be affected by an overstated immigrant-native elasticity.

<sup>18</sup>Across years, the average education-experience-race cell is constructed from 115 observations for blacks and 1,868 observations for whites.

<sup>19</sup>Boustan (2009), studying the 1940–1970 period, obtains somewhat smaller estimates, which I can closely replicate by adopting her sample restrictions (including foreign-born and other nonwhite men, excluding the self-employed, those enrolled in school, and those living in group quarters) and experience imputation scheme. As I show below, however, the estimated effects of Southern immigration are not sensitive to small differences in the assumed black-white elasticity of substitution; the differences between our estimates appear to be driven by differences in the weights used to compute wage changes.

In contrast, Borjas et al. (2010) find no evidence of imperfect substitution in the period spanning 1960–2000. The pattern of elasticities that increase with time is consistent with black-white relative progress during the 20th Century.

<sup>20</sup>Ottaviano and Peri (2012, Appendix A1) sketch a proof. Here I provide another for the simplified case where output is produced according to  $L = \left( \sum_{i \in \{1, \dots, I\}} \theta_i L_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$  and its price is normalized to one. Differentiating the marginal product condition for profit maximization gives  $\partial \log w_i / \partial \log L_j = (1/\sigma)(\partial \log L / \partial L_j)$ . Let  $c \equiv c(L(L_1, \dots, L_I), w)$  be the minimal cost of producing  $L(L_1, \dots, L_I)$  units of output when wages are  $w$ . Because production is constant returns to scale,  $c = c(1, w)L(L_1, \dots, L_I)$ . Since output is the numeraire,  $c(1, w) = 1$  in competitive equilibrium. Taking logs,  $\partial \log c / \partial \log L_j = \partial \log L / \partial \log L_j$ . We also have that  $\partial \log c / \partial \log L_j = (\partial c / \partial L_j) / (c / L_j)$ . Since profit maximization implies that  $\partial c / \partial L_j = (\partial c / \partial L)(\partial L / \partial L_j) = w_j$  in a competitive equilibrium where output is the numeraire,  $\partial \log c / \partial \log L_j = w_j L_j / c = s_j$ . Thus,  $\partial \log w_i / \partial \log L_i = (1/\sigma)(\partial \log L / \partial L_j) = s_j / \sigma$ .

<sup>21</sup>Also note that the long-run effects of immigration are necessarily distributional, at least if capital is perfectly mobile—because production is CRS, capital investments will completely offset changes to the productivity of (and hence the remuneration paid to) aggregate labor, leaving the average wage unchanged.

<sup>22</sup>However, since I assume the larger black-white elasticities implied by the coefficients in Tables 12 and 14, these calculations are conservative.

<sup>23</sup>These estimates are not directly comparable since my wage regressions relate wages to immigrant stocks while my simulations relate changes in wages to immigrant flows. However, since few Southern blacks lived in the North before 1915, the 1950 stock of immigrants is also roughly the 1915–1950 flow, providing a basis for comparison.

## Appendix 1

### Local labor markets appendix

#### *Sample and variable construction*

My local labor markets analysis draws on 1 % extracts of the 1940 and 1950 US Censuses, provided by IPUMS (Ruggles et al. 2015). Note that the 1950 Census only asked a subset of households wage information, so the regression samples for that decade are smaller. I use the Eichenlaub et al. (2010) definition of the “South”, classifying Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, and Virginia as Southern and all other states as Northern.

My sample consists of US-born black and white men, aged 16–64, who reported nonzero earnings and lived in the North at the time of enumeration. I define labor markets as metropolitan statistical areas. Although the 1940 Census did not use an MSA concept,

IPUMS identifies areas in the 1940 Census that fall within 1950 MSA boundaries and are sufficiently populated to meet confidentiality requirements (Ruggles et al. 2015). Blacks living in the North were clustered into a relatively small number of MSAs. To eliminate the possibility of influential outlier MSAs consisting of only a few black men, I drop MSAs with fewer than 10 black men meeting the other sample criteria. Appendix 3: Table 18 lists the metropolitan areas used in the black regression samples for each year.

I use the CPI weights supplied by IPUMS to inflate all wage variables to 1999 dollars. Annual wages consist of all wage income during the year and weekly wages are annual wages divided by weeks worked.

Some of the covariates that I use are MSA-level averages. In constructing these variables, I weight each observation by the person weights supplied by IPUMS.

**Misspecified lagged dependent variables and fixed-effects models**

Guryan (2001) shows that, when  $x$  is a serially uncorrelated treatment indicator, estimates from lagged dependent variables and fixed effects specifications bound the treatment effect when the data-generating process is unknown. I extend his argument to the case of serially correlated treatments. I also compare the biases of models with and without lagged dependent variables when the data-generating processes exhibits permanent unobserved heterogeneity. I focus on the simplified case where the treatment effect is zero under the null hypothesis in order to emphasize the bias due to misspecification. I assume that all variables are area-level means with individual covariates (and other area-level covariates) partialled out.

Suppose that  $x$  has no causal effect, so that the value of  $y$  in area  $j$  evolves according to the data-generating process

$$y_{jt} = a_j + e_{jt},$$

with  $E(e_{jt}|a_j) = 0$ , implying that  $y_{jt} = y_{jt-1} + e_{jt} - e_{jt-1}$ .

If the regression is misspecified as including the lagged dependent variable, i.e.,  $y_{jt} = \alpha y_{jt-1} + \beta x_{jt} + u_{jt}$ , the probability limit of  $\hat{\beta}$  is

$$\begin{aligned} \beta_{LDV}^* &= \frac{Cov(y_{jt}, \tilde{x}_{jt})}{Var(\tilde{x}_{jt})} = \frac{Cov(y_{jt-1} + e_{jt} - e_{jt-1}, \tilde{x}_{jt})}{Var(\tilde{x}_{jt})} \\ &= \frac{Cov(e_{jt} - e_{jt-1}, x_{jt} - \gamma_1(a_j + e_{jt-1}))}{Var(\tilde{x}_{jt})} = \gamma_1 \frac{\sigma_e^2}{Var(\tilde{x}_{jt})} \\ &= \frac{Cov(x_{jt}, a_j)}{Var(y_{jt-1})} \cdot \frac{\sigma_e^2}{Var(\tilde{x}_{jt})}, \end{aligned}$$

where  $\tilde{x}_{jt} = x_{jt} - \gamma_0 - \gamma_1 y_{jt-1}$  is the residual from a population regression of  $x_{jt}$  on  $y_{jt-1}$  and  $\sigma_e^2$  is the variance of  $e_{jt}$ , which is assumed to be iid across  $j$  and  $t$ . If  $x$  is positively correlated with  $a$ , then  $\beta_{LDV}^* > 0$ .

Now suppose the data-generating process is

$$y_{jt} = \delta y_{jt-1} + e_{jt}$$

with  $\delta < 1$ , but the model is misspecified as  $y_{jt} = \beta x_{jt} + a_j + u_{jt}$ . In addition, suppose that over short periods of time the process by which  $x_{jt}$  evolves can be described by  $x_{jt} = \lambda x_{jt-1} + r_{jt}$ , with  $\lambda > 1$  (to reflect the accumulation of immigrants over time) and  $r$  is white noise. In this case, the probability limit of  $\hat{\beta}$  is

$$\begin{aligned}
\beta_{FD}^* &= \frac{Cov(y_{jt} - y_{jt-1}, x_{jt} - x_{jt-t})}{Var(x_{jt} - x_{jt-t})} \\
&= \frac{Cov(y_{jt}, x_{jt}) - Cov(y_{jt}, x_{jt-1}) - Cov(y_{jt-1}, x_{jt}) + Cov(y_{jt-1}, x_{jt-1})}{Var(x_{jt} - x_{jt-1})} \\
&= \frac{(\delta\lambda)^{t-1}(\lambda\delta - \lambda - \delta + 1)\sigma_{xa}}{Var(x_{jt} - x_{jt-1})} = \frac{(\delta\lambda)^{t-1}[(\lambda - 1)(\delta - 1)]\sigma_{xa}}{Var(x_{jt} - x_{jt-1})} < 0,
\end{aligned}$$

where  $\sigma_{xa} = Cov(x_0, y_{j0})$  is positive by assumption.

We thus have that  $\beta_{FD}^* < \beta = 0 < \beta_{LDV}^*$ .

A similar result allows us to compare the OLS and lagged dependent variable estimators when there is permanent heterogeneity. If the data-generating process is  $y_{jt} = a_j + e_{jt}$  but the regression is specified as  $y_{jt} = \beta x_{jt} + u_{jt}$ , the population regression coefficient will be

$$\beta_{OLS}^* = \frac{Cov(y_{jt}, x_{jt})}{Var(x_{jt})} = \frac{Cov(a_j, x_{jt})}{Var(x_{jt})} > 0.$$

Thus, the lagged dependent variable estimate will be less biased than standard OLS if

$$\frac{Var(x_{jt})}{Var(\tilde{x}_{jt})} \cdot \frac{\sigma_e^2}{Var(y_{jt-1})} < 1.$$

Note that, since  $\sigma_e^2/Var(y_{jt-1}) \leq 1$  by definition, this condition is testable. In my dataset,  $Var(x_{jt})/Var(\tilde{x}_{jt}) \approx 1$ .

## Appendix 2

### National labor market appendix

#### Sample and variable construction

The education categories that I use are less than fifth grade (less than 5 years of schooling), fifth-to-eighth grade (between 5 and 8 years), some high school (between 9 and 11 years), high school (12 years), and greater than high school (greater than 12 years). I assume that those with less than a fifth grade education spend 2.5 years in school, those with a fifth-to-eighth grade education spend 7.5, those with some high school spend 10, those with a high school degree spend 12, and those with more than a high school degree spend 13.5. Imputed experience is age, minus six, minus years spent in school. I then assign individuals to one of eight 5-year experience groups (one to five, six to ten, etc.) and drop individuals with less than one or more than 40 years of experience.

With my sample restrictions, there are no individuals in the sample in the lowest education-experience group. I then have 39 education experience groups for a total of 234 education-experience-race-year observations (which include natives and immigrants) or 312 observations when 1970 is included in the sample. My measure of labor supply within education-experience-race cells for each year is the sum of the Census sampling weights provided by IPUMS and my wage measure is the sampling-weighted average annual wage.

#### Implementation details

The aggregate production function is  $Y = AL^\alpha K^{1-\alpha}$  where  $L$  is successively CES-nested as described in the main text. The competitive wage for a member of group  $(e, x, r, i)$  is

$$\begin{aligned} \log w_{exri} = & \log \left[ A\alpha \left( \frac{K}{L} \right)^{1-\alpha} \right] + \frac{1}{\sigma_e} \log L + \log \theta_e + \left( \frac{1}{\sigma_x} - \frac{1}{\sigma_e} \right) \log L_e \\ & + \log \theta_{ex} + \left( \frac{1}{\sigma_r} - \frac{1}{\sigma_x} \right) \log L_{ex} + \log \theta_{exr} + \left( \frac{1}{\sigma_i} - \frac{1}{\sigma_r} \right) \log L_{exr} \\ & + \log \theta_{exri} - \frac{1}{\sigma_i} \log L_{exri}, \end{aligned} \tag{5}$$

which implies (1). Similarly, wages paid to members of education-experience-race groups satisfy

$$\log \left( \frac{\bar{w}_{exwt}}{\bar{w}_{exbt}} \right) = \log \left( \frac{\theta_{exwt}}{\theta_{exbt}} \right) - \frac{1}{\sigma_r} \log \left( \frac{L_{exwt}}{L_{exbt}} \right),$$

which motivates specification (3) with  $\bar{w}_{exrt} = w_{exrnt} \cdot (L_{exrnt}/L_{exrt}) + w_{exrst} \cdot (L_{exrst}/L_{exrt})$  for  $r \in \{b, w\}$ .<sup>24</sup>

The model implies that

$$\log \bar{w}_{ext} = \log \left[ A_t \alpha \left( \frac{K_t}{L_t} \right)^{1-\alpha} \right] + \frac{1}{\sigma_e} \log L_t + \log \theta_{et} + \left( \frac{1}{\sigma_x} - \frac{1}{\sigma_e} \right) \log L_{et} + \log \theta_{ex} - \frac{1}{\sigma_x} \log L_{ext},$$

so that  $\sigma_x$  can be recovered from the regression specification

$$\log \bar{w}_{ext} = \lambda_{et} + \lambda_{ex} - \frac{1}{\sigma_x} \log \hat{L}_{ext} + u_{ext}. \tag{6}$$

To construct the  $\hat{L}_{ext}$  and  $\bar{w}_{ext}$ , I assume that  $\theta_{exw} + \theta_{exb} = 1$  and estimate

$$\begin{aligned} \hat{\theta}_{wext} = 1 - \hat{\theta}_{bext} &= \frac{\exp(\hat{\lambda}_{ex}) \exp(\hat{\lambda}_{et}) \exp(\hat{\lambda}_{xt})}{1 + \exp(\hat{\lambda}_{ex}) \exp(\hat{\lambda}_{et}) \exp(\hat{\lambda}_{xt})}, \\ \hat{L}_{ext} &= \left( \hat{\theta}_{exwt} \hat{L}_{exwt}^{\frac{\hat{\sigma}_r}{\hat{\sigma}_r-1}} + \hat{\theta}_{exbt} \hat{L}_{exbt}^{\frac{\hat{\sigma}_r}{\hat{\sigma}_r-1}} \right)^{\frac{\hat{\sigma}_r-1}{\hat{\sigma}_r}}, \end{aligned}$$

and  $\bar{w}_{ext} = \bar{w}_{exwt} \cdot (\hat{L}_{exwt}/\hat{L}_{ext}) + \bar{w}_{exbt} \cdot (\hat{L}_{exbt}/\hat{L}_{ext})$ , where the fixed effects and elasticity estimates are taken from the OLS estimates of (2) applied to the entire sample.

Finally, the model implies that

$$\log \bar{w}_{et} = \log \left[ \alpha A_t \left( \frac{K_t}{L_t} \right)^{1-\alpha} \right] + \frac{1}{\sigma_e} \log L_t + \log \theta_{et} - \frac{1}{\sigma_e} \log L_{et},$$

which suggests that  $\sigma_e$  can be recovered from estimates of

$$\log \bar{w}_{et} = \lambda_t + \sum_e \kappa_e \cdot t - \frac{1}{\sigma_e} \log \hat{L}_{et} + u_{et}. \tag{7}$$

Following Borjas (2003) and Ottaviano and Peri (2012), I assume that the  $\theta_{ex}$  are time invariant (in order to recover them separately from the aggregate time terms in (6)) with  $\sum_x \theta_{ex} = 1$  and estimate

$$\hat{\theta}_{ex} = \frac{\exp(\hat{\lambda}_{ex})}{\sum_x \exp(\hat{\lambda}_{ex})}, \quad \hat{L}_e = \left( \sum_x \hat{\theta}_{ex} \hat{L}_{ext}^{\frac{\hat{\sigma}_x-1}{\hat{\sigma}_x}} \right)^{\frac{\hat{\sigma}_x}{\hat{\sigma}_x-1}},$$

and  $\bar{w}_{et} = \sum_x \bar{w}_{ext} (\hat{L}_{ext}/\hat{L}_{et})$ , where the fixed effects and elasticities are taken from the IV estimates of (6).

**Appendix 3**  
**Tables**

**Table 17** Southern immigration and metro-average annual wages

Specification	OLS				IV				
	1940		1950		1940		1950		
	Black	White	Black	White	Black	White	Black	White	
(1)	Prop. Southern	0.265 (0.19)	0.449** (0.19)	-0.485* (0.25)	0.450*** (0.15)	-0.0345 (0.37)	0.294 (0.24)	-0.782*** (0.24)	0.483*** (0.18)
	Observations	27	93	23	100	26	80	22	80
(2)	Prop. Southern	0.0215 (0.45)	0.369** (0.15)	-0.765 (0.48)	0.449*** (0.15)	-0.837*** (0.27)	0.286 (0.20)	-0.796* (0.43)	0.454** (0.23)
	Observations	27	87	23	72	26	75	22	62
(3)	Prop. Southern			-0.977*** (0.33)	0.172 (0.13)			-1.488*** (0.35)	0.251 (0.22)
	Observations			23	87			22	75
(4)	Prop. Southern			-1.243** (0.51)	0.251** (0.10)			-1.783*** (0.46)	0.294 (0.27)
	Observations			23	67			22	59

Notes: Dependent variable is metro-average residuals from a regression of log annual wages on indicators for educational attainment and age. Specification (2) includes white and black metro-level percent employed in manufacturing, percent farming, and average years of education; specification (3) includes the metro level average dependent variable, lagged one decade; specification (4) includes metro-level average variables and a lagged mean dependent variable. Regressions weighted by the number of observations used to calculate metro-level covariates. Robust standard errors reported in parentheses  
 \*\*\*, \*\*, and \* denote significance at the 1, 5, and 10 % levels, respectively

**Table 18** Metro areas used in black regressions

Metro	1940					1950				
	Prop. Southern			# of natives		Prop. Southern			# of natives	
	All	Black	White	Black	White	All	Black	White	Black	White
Atlantic City, NJ	0.094	0.085	0.009	23	181	0.030	0.022	0.008	13	54
Baltimore, MD	0.149	0.083	0.066	237	2062	0.213	0.116	0.097	85	735
Boston, MA-NH	0.015	0.007	0.007	41	4582	0.032	0.012	0.020	18	1725
Charleston, WV	0.161	0.057	0.104	26	543	0.113	0.059	0.054	14	204
Chicago, IL	0.095	0.062	0.032	155	9003	0.149	0.102	0.047	105	3308
Cincinnati-Hamilton, OH/KY/IN	0.266	0.082	0.184	30	996	0.281	0.085	0.196	12	390
Cleveland, OH	0.094	0.065	0.029	37	2368	0.113	0.085	0.028	29	877
Columbus, OH	0.106	0.061	0.045	38	799	0.098	0.043	0.055	19	349
Dayton-Springfield, OH	0.137	0.056	0.082	37	872	0.259	0.106	0.153	11	323
Des Moines, IA	0.027	0.005	0.023	11	418					
Detroit, MI	0.157	0.073	0.085	62	4226	0.245	0.135	0.110	51	1616
Indianapolis, IN	0.176	0.077	0.099	39	911	0.167	0.063	0.104	18	311
Kansas City, MO-KS	0.139	0.057	0.083	96	1311	0.179	0.081	0.098	21	492
Los Angeles-Long Beach, CA	0.171	0.024	0.147	44	4737	0.200	0.049	0.151	29	2246
Minneapolis-St. Paul, MN	0.020	0.007	0.013	15	1898					
New Haven-Meriden, CT	0.018	0.011	0.007	15	977					
New York-Northeastern NJ	0.066	0.045	0.020	427	19203	0.083	0.058	0.025	235	6835
Omaha, NE/IA	0.061	0.034	0.027	11	545	0.077	0.027	0.050	10	184
Philadelphia, PA/NJ	0.087	0.064	0.024	263	5729	0.128	0.094	0.033	136	2067
Pittsburgh, PA	0.061	0.044	0.017	79	4007	0.064	0.050	0.015	39	1436
St. Louis, MO-IL	0.131	0.070	0.061	97	2805	0.168	0.096	0.072	45	978
San Francisco-Oakland-Vallejo, CA	0.100	0.011	0.089	18	2719	0.202	0.059	0.143	13	1204
Seattle-Everett, WA						0.129	0.030	0.099	10	458
Topeka, KS	0.068	0.029	0.039	11	182					
Trenton, NJ						0.084	0.073	0.011	12	147
Washington, DC/MD/VA	0.354	0.151	0.202	232	1128	0.361	0.148	0.213	106	459
Wheeling, WV/OH	0.049	0.028	0.022	14	692					
Wilmington, DE/NJ/MD	0.050	0.017	0.032	37	353					
Youngstown-Warren, OH-PA	0.070	0.045	0.025	12	934	0.063	0.052	0.011	12	368



**Table 19** First-stage regressions

Dependent variable	Instrument(s)	Individual covariates				Individual and metro-level covariates			
		1940		1950		1940		1950	
		Black	White	Black	White	Black	White	Black	White
Prop. Southern	Predicted Prop. Southern	0.544*** (0.0724)	0.514*** (0.0739)	0.453*** (0.0778)	0.372*** (0.0852)	0.559*** (0.0647)	0.532*** (0.0618)	0.289*** (0.0803)	0.358*** (0.0697)
	Observations	2081	104,029	1029	39,219	2081	101,339	1029	36,413
	R-squared	0.852	0.606	0.628	0.272	0.932	0.772	0.836	0.570
	Clusters	26	80	22	80	26	75	22	62
Prop. Southern black	Predicted Prop. Southern black	0.322*** (0.0829)	0.471*** (0.0918)	0.300*** (0.0520)	0.376*** (0.0733)	0.313*** (0.0985)	0.469*** (0.0653)	0.233*** (0.0697)	0.329*** (0.0546)
	Predicted Prop. Southern white	0.115 (0.0773)	-0.0522 (0.0518)	0.0676 (0.0411)	0.0165 (0.0569)	0.0863 (0.0538)	-0.00589 (0.0563)	0.0345 (0.0401)	0.0206 (0.0523)
	Observations	2081	104,029	1029	39,219	2081	101,339	1029	36,413
	R-squared	0.834	0.732	0.555	0.452	0.933	0.833	0.877	0.677
	Clusters	26	80	22	80	26	75	22	62
Prop. Southern white	Predicted Prop. Southern black	-0.197 (0.139)	-0.289 (0.175)	-0.257 (0.158)	-0.413** (0.173)	-0.616*** (0.182)	-0.209 (0.126)	-0.416*** (0.142)	-0.279** (0.105)
	Predicted Prop. Southern white	0.748*** (0.110)	0.874*** (0.187)	0.725*** (0.109)	0.807*** (0.159)	1.018*** (0.131)	0.779*** (0.127)	0.685*** (0.102)	0.653*** (0.115)
	Observations	2081	104,029	1029	39,219	2081	101,339	1029	36,413
	R-squared	0.879	0.717	0.779	0.594	0.957	0.799	0.893	0.720
	Clusters	26	80	22	80	26	75	22	62

Notes: All specifications include indicators for age and education; metro-level covariates include white and black percent employed in manufacturing, percent farming, and average years of education. Standard errors, clustered by metro area, reported in parentheses

\*\*\*, \*\*, and \* denote significance at the 1, 5, and 10 % levels, respectively

**Table 20** Weekly wages and immigration

Specification		OLS				IV			
		1940		1950		1940		1950	
		Black	White	Black	White	Black	White	Black	White
(1)	Prop. Southern	-0.301 (0.197)	0.164 (0.307)	-0.415 (0.294)	0.396** (0.153)	-0.579 (0.362)	-0.259 (0.333)	-0.940** (0.466)	0.120 (0.241)
	Observations	2106	107,104	1029	40,751	2080	103,986	1015	38,839
	Clusters	27	93	23	100	26	80	22	80
(2)	Prop. Southern	-0.455 (0.323)	0.324 (0.256)	-0.582 (0.422)	0.379** (0.181)	-0.832*** (0.282)	0.0717 (0.286)	-0.919 (0.592)	0.0892 (0.234)
	Observations	2106	104,167	1029	37,065	2080	101,299	1015	36,055
	Clusters	27	87	23	72	26	75	22	62
(3)	Prop. Southern			-0.929*** (0.257)	0.164* (0.0860)			-1.221*** (0.258)	0.159 (0.185)
	Observations			1029	38,933			1015	26,290
	Clusters			23	87			22	22
(4)	Prop. Southern			-1.129*** (0.265)	0.145* (0.0767)			-1.256*** (0.425)	-0.0631 (0.177)
	Observations			1029	36,114			1015	26,290
	Clusters			23	67			22	22

Notes: All specifications include indicators for age and education; specification (2) includes white and black metro-level percent employed in manufacturing, percent farming, and average years of education; specification (3) includes the metro level average dependent variable, lagged one decade; specification (4) includes metro-level average variables and a lagged mean dependent variable. Regressions weighted by the number of observations used to calculate metro-level covariates. Standard errors, clustered by metro area, reported in parentheses  
 \*\*\*, \*\*, and \* denote significance at the 1, 5, and 10 % levels, respectively

**Table 21** Weekly wages and race-specific immigration

Specification		OLS				IV			
		1940		1950		1940		1950	
		Black	White	Black	White	Black	White	Black	White
(1)	Prop. Southern black	-1.066 (0.722)	1.328** (0.523)	-1.294 (1.076)	1.091*** (0.241)	-1.501** (0.696)	0.164 (0.843)	-3.607*** (1.197)	0.262 (0.444)
	Prop. Southern white	0.0966 (0.392)	-0.212 (0.222)	0.0705 (0.332)	0.0776 (0.131)	-0.0224 (0.470)	-0.526** (0.257)	0.718 (0.564)	-0.0536 (0.141)
	Observations	2106	107,104	1029	40,751	2080	103,986	1015	38,839
	Clusters	27	93	23	100	26	80	22	80
(2)	Prop. Southern black	-0.938 (1.236)	1.198*** (0.427)	-2.206 (1.338)	1.007*** (0.336)	-2.887** (1.316)	0.569 (0.525)	-2.626** (1.289)	0.178 (0.435)
	Prop. Southern white	-0.273 (0.617)	-0.0875 (0.247)	0.0998 (0.394)	0.0364 (0.173)	0.0964 (0.638)	-0.272 (0.290)	0.443 (0.412)	0.00247 (0.233)
	Observations	2106	104,167	1029	37,065	2080	101,299	1015	36,055
	Clusters	27	87	23	72	26	75	22	62
(3)	Prop. Southern black			-1.279** (0.616)	0.685*** (0.180)			-2.050*** (0.633)	0.374* (0.200)
	Prop. Southern white			-0.693 (0.451)	-0.0290 (0.0631)			-0.583 (0.571)	-0.0911 (0.115)
	Observations			1029	38,933			1015	37,712
	Clusters			23	87			22	75
(4)	Prop. Southern black			-3.193*** (0.738)	0.466** (0.213)			-2.518*** (0.811)	0.189 (0.210)
	Prop. Southern white			-0.374 (0.329)	-0.0113 (0.105)			-0.128 (0.374)	0.00415 (0.153)
	Observations			1029	36,114			1015	35,295
	Clusters			23	67			22	59

Notes: All specifications include indicators for age and education; specification (2) includes white and black metro-level percent employed in manufacturing, percent farming, and average years of education; specification (3) includes the metro level average dependent variable, lagged one decade; specification (4) includes metro-level average variables and a lagged mean dependent variable. Regressions weighted by the number of observations used to calculate metro-level covariates. Standard errors, clustered by metro area, reported in parentheses

\*\*\*\*, \*\*\*, and \*\* denote significance at the 1, 5, and 10 % levels, respectively

**Table 22** Changes in South-North migration

Education group	Experience group	Proportional change (1940–1960)		Proportional change (1940–1970)	
		White	Black	White	Black
Less than 5th grade	6–10	0.10	−0.50	−0.15	−1.00
	11–15	−0.02	0.20	−0.48	−0.75
	16–20	0.58	0.24	−0.19	−0.65
	21–25	0.58	0.11	−0.23	−0.71
	26–30	0.90	0.08	0.33	−0.69
	31–35	0.25	0.14	0.30	−0.58
	36–40	0.98	0.21	1.21	−0.40
5th–8th grade	1–5	2.14	3.05	2.63	5.69
	6–10	0.04	0.13	−0.39	−0.60
	11–15	0.10	0.01	−0.37	−0.61
	16–20	0.25	0.09	−0.34	−0.65
	21–25	0.09	0.20	−0.24	−0.47
	26–30	0.15	0.19	−0.02	−0.37
	31–35	0.62	0.81	0.63	0.33
Some high school	36–40	0.83	1.19	0.72	0.78
	1–5	3.06	1.66	4.62	2.33
	6–10	0.83	1.92	0.70	1.36
	11–15	0.63	1.82	0.46	1.59
	16–20	0.88	2.72	0.59	2.41
	21–25	1.50	3.60	1.22	3.11
	26–30	2.11	4.79	2.58	6.97
High school degree	31–35	2.02	4.49	3.60	9.39
	36–40	3.87	5.48	5.94	11.88
	1–5	1.95	3.04	2.56	4.38
	6–10	1.18	2.29	2.22	4.47
	11–15	1.09	4.62	1.87	11.38
	16–20	1.92	4.34	3.09	9.02
	21–25	3.32	5.32	4.21	6.46
Greater than high school	26–30	3.28	6.31	5.26	16.27
	31–35	2.63	8.73	6.42	14.33
	36–40	2.75	4.20	8.65	8.27
	1–5	1.72	2.64	5.74	6.52
	6–10	1.94	3.54	3.81	8.23
	11–15	1.84	4.60	3.00	8.50
	16–20	2.24	2.71	3.50	6.20
Greater than high school	21–25	1.78	2.34	3.12	4.29
	26–30	1.31	2.19	4.74	4.89
	31–35	1.63	1.76	4.36	4.55
	36–40	1.90	2.35	5.78	4.18

**Competing interests**

The IZA Journal of Migration is committed to the IZA Guiding Principles of Research Integrity. The author declares that he has observed these principles.

**Acknowledgements**

I thank Lowell Taylor, Brian Kovak, George-Levi Gayle, Leah Platt Boustan, and seminar participants at Carnegie Mellon University, IZA, the Southern Economics Association meetings, and the University of Mississippi for helpful comments and suggestions.

I would also like to thank an anonymous referee and the editor for useful remarks.

Responsible editor: Denis Fougère

Received: 6 April 2016 Accepted: 13 July 2016

Published online: 28 November 2016

## References

- Altonji JG, Card D. The effects of immigration on the labor market outcomes of less-skilled natives In: Abowd JM, Freeman RB, editors. *Immigration, Trade and the Labor Market*. Chicago and London: University of Chicago Press; 1991. p. 201–34.
- Aydemir A, Borjas GJ. Attenuation bias in measuring the wage impact of immigration. *J Labor Econ*. 2011;29(1):69–112.
- Bartel A. Where do the new U.S. immigrants live? *J Labor Econ*. 1989;7:371–91.
- Becker G. *The Economics of Discrimination*, 2nd edn. Chicago and London: University of Chicago Press; 1971.
- Borjas GJ. The labor demand curve is downward sloping: Rexamining the impact of immigration on the local labor market. *Q J Econ*. 2003;118(4):1335–74.
- Borjas GJ. Native internal migration and the labor market impact of immigration. *J Hum Res*. 2006;41(2):221–58.
- Borjas GJ, Grogger J, Hanson GH. Immigration and the economic status of African-American men. *Economica*. 2010;77:255–82.
- Boustan LP. Competition in the promised land: Black migration and racial wage convergence in the North, 1940–1970. *J Econ Hist*. 2009;69(3):756–83.
- Boustan LP, Fishback PV, Kantor S. The effect of internal migration on local labor markets: American cities during the Great Depression. *J Labor Econ*. 2010;28(4):719–46.
- Card D. The impact of the Mariel Boatlift on the Miami labor market. *Ind Labor Relat Rev*. 1990;43(2):245–57.
- Card D. Immigrant inflows, native outflows and the local labor market impacts of higher immigration. *J Labor Econ*. 2001;19:22–64.
- Card D. Immigration and inequality. *Am Econ Rev (Papers and Proceedings)*. 2009;99(2):1–21.
- Card D, DiNardo J. Do immigrant inflows lead to native outflows? *Am Econ Rev (Papers and Proceedings)*. 2000;90(2):360–7.
- Card D, Lemieux T. Can falling supply explain the rising returns to college for younger men? A cohort based analysis. *Q J Econ*. 2001;116(2):705–46.
- Donohue III JJ, Heckman J. Continuous versus episodic change: The impact of civil rights policy on the economic status of blacks. *J Econ Lit*. 1991;29(4):1603–43.
- Dumont M, Rayp G, Thas O, Wileme P. Correcting standard errors in two-stage estimation procedures with generated regressands. *Oxf Bull Econ Stat*. 2005;67(3):421–33.
- Dustmann C, Schonberg U, Stuhler J. Labor supply shocks, native wages, and the adjustment of local employment. Manuscript, University College London; 2016.
- Eichenlaub SC, Tolnay SE, Alexander JT. Moving out but not up: Economic outcomes in the Great Migration. *Am Soc Rev*. 2010;75:101–25.
- Guryan J. Desegregation and black dropout rates. 2001. Working paper no. 8345. National Bureau of Economic Research. <http://www.nber.org/papers/w8345>.
- Lang KM, Gottschalk P. The loss in efficiency from using grouped data to estimate coefficients of group level variables. *Comput Econ*. 1996;9(4):355–61.
- Massey DS, Denton NA. *American Apartheid: Segregation and the making of the underclass*. Cambridge, Mass: Harvard University Press; 1993.
- Ottaviano GIP, Peri G. Rethinking the effect of immigration on wages. *J Eur Econ Assoc*. 2012;10(1):152–97.
- Peri G, Sparber C. Assessing inherent model bias: An application to native displacement in response to immigration. *J Urban Econ*. 2011;69(1):82–91.
- Ruggles S, Genadek K, Goeken R, Grover J, Sobek M. *Integrated Public Use Microdata Series: Version 6.0* [Machine-readable database]. Minneapolis: University of Minnesota; 2015. <https://usa.ipums.org/usa/>.
- Smith JP, Welch FR. Race differences in earnings: a survey and new evidence. Santa Monica, CA: RAND Corporation; 1978. <http://www.rand.org/pubs/reports/R2295.html>.
- Smith JP, Welch FR. Black economic progress after Myrdal. *J Econ Lit*. 1989;27(2):519–64.
- Tolnay SE. The African American “Great Migration” and beyond. *Ann Rev Sociol*. 2003;29:209–32.
- Tolnay SE, White KJC, Crowder KD, Adelman RM. Distances traveled during the Great Migration: an analysis of racial differences among male migrants. *Soc Sci Hist*. 2005;29(4):523–48.
- Welch FR. Effects of cohort size on earnings: the baby boom babies’ financial boost. *J Pol Econ*. 1979;87:65–97.
- White MJ, Lindstrom DP. Internal migration. In: Poston DL, Micklin M, editors. *Handbook of Population*. Boston MA: Springer; 2005. p. 311–346.