Black Networks After Emancipation: Evidence from Reconstruction and the Great Migration *

Kenneth Chay† Kaivan Munshi‡

March 2013

Abstract

We find that southern blacks responded collectively to political and economic opportunities after Emancipation, but only in places where specific conditions were satisfied. Blacks would have lived and worked in close proximity to each other in counties where labor intensive plantation crops were grown. Spatial proximity would have resulted in more connected social networks. These connected networks would have supported the formation of larger coalitions of black activists during Reconstruction and larger coalitions of black workers moving together to northern cities during the Great Migration. Our theoretical model places additional structure on this relationship. Coalitions will not form and there will be no association between network connectedness and the outcomes of interest – political participation and migration – up to a threshold, followed by a positive association above the threshold. Voting and migration patterns across counties are consistent with the theory - there is no association with our crop-based measure of network connectedness up to a threshold point at which a steep, monotonic relationship begins. This finding is robust to rigorous testing, and these tests show that competing hypotheses do not exhibit similar nonlinear patterns. Blacks from southern counties where plantation crops were grown accounted for a disproportionate share of northern migrants, and these migrants appear to have benefited from network externalities, as they moved to the same destination cities.

Keywords. Networks. Social capital. Coalition formation. African-American history. Slavery. Reconstruction. Great Migration.
JEL. D85. J62. L14. L22.

*We are grateful to Jan Eeckhout, Seth Rockman, and Yuya Sasaki for many helpful discussions. Bob Margo, Gavin Wright and numerous seminar participants provided comments that substantially improved the paper. Jacob Goldston, Florencia Borrescio Higa, Shiva Koohi, and, especially, Bruno Gasperini provided outstanding research support over the course of this project. Dan Black, Seth Sanders, Evan Taylor, and Lowell Taylor graciously provided us with the Mississippi migration data and Jeremy Atack and Bob Margo provided us with the railroads data. Financial support from the National Science Foundation through grant SES-0617847 is gratefully acknowledged. We are responsible for any errors that may remain.

†Brown University and NBER
‡Brown University and NBER

1 Introduction

Were African-Americans able to overcome centuries of social dislocation and form viable communities once they were free? This question has long been debated by social historians, and is relevant both for contemporary social policy and for understanding the process of social capital formation. The traditional view was that slavery, through forced separation and by restricting social interaction, permanently undermined the black community (Du Bois 1908, Frazier 1939, Stampp 1956). This was replaced by a revisionist history that documented a stable, vibrant African-American family and community, both during and after slavery (Blassingame 1972, Genovese 1974, Gutman 1976). More recently, Fogel (1989) and Kolchin (1993) have taken a position between the traditional and the revisionist view; while other social scientists have brought the literature around full circle by asserting that "[s]lavery was, in fact, a social system *designed* to destroy social capital among slaves" (Putnam 2000: 294).[1]

Despite the importance of and continuing interest in black social capital after slavery, there has been virtually no quantitative investigation in this area. We cannot examine the impact of slavery on social capital – i.e. the social capital that would have prevailed in the absence of slavery. What we can study is the equally important question of whether (and where) blacks formed coalitions to achieve common objectives soon after slavery ended. In the decades following the Civil War, two significant opportunities arose for southern blacks to work together. First, blacks were able to vote and elect their own leaders during and just after Reconstruction, 1870-1890. Second, blacks were able to leave the South and find jobs in northern cities during the Great Migration, 1916-1930. We find that southern blacks did organize themselves in both these events, but only in places where specific historical preconditions were satisfied.

The point of departure for our analysis is the observation that black population densities varied substantially across counties in the South, depending on the the crops that were grown in the local area. Where labor intensive crops such as tobacco, cotton, rice, and sugarcane were grown, blacks lived and worked in close proximity to each other. Where crops such as wheat and corn were grown, blacks were dispersed more widely. Spatial proximity would have resulted in more connected social networks. We show theoretically that these connected networks would have supported the formation of larger coalitions of black political activists during Reconstruction and larger coalitions of black workers moving together to northern cities during the Great Migration. This, in turn, would have given rise to greater overall political participation and migration.

[1]Putnam goes on from there to argue that current differences in social capital across U.S. states can be traced back to historical slavery.

network connectedness $\rightarrow$ *coalition size* $\rightarrow$ *political participation and migration.*

While a positive relationship between network connectedness and particular outcomes during Reconstruction and the Great Migration is consistent with the presence of underlying (unobserved) black coalitions, other explanations are available. For example, racial conflict could have been greater in counties where plantation crops were grown, resulting in a larger black voter-turnout during Reconstruction and greater movement to northern cities during the Great Migration. Alternatively, adverse economic conditions in these counties could have encouraged greater political participation and migration. Our strategy to identify the presence of underlying coalitions takes advantage of an additional prediction of the theory, which is that coalitions will only form above a threshold level of network connectedness. There should thus be no association between the outcomes of interest – political participation and migration – and network connectedness up to a threshold and a positive association thereafter. This specific nonlinearity, characterized by a slope discontinuity at a threshold, is central to our identification strategy.

Apart from its contribution to quantitative African-American history, this paper links the emerging literature on social networks in economics to the extant literature on coalition formation. The social networks literature has devoted much attention to the level of trust or cooperation that can be supported between two nodes in a network. The general result is that connected networks support cooperation because deviators can be punished more severely (Granovetter 1985, Coleman 1988, Karlan et al. 2009, Jackson et al. 2012).[2] More recently, some papers have gone beyond the analysis of bilateral links to study public goods provision in networks (Ballester et al. 2006, Elliott and Golub 2013). The main result from these papers is that the level of effort devoted by each individual will depend on the benefit he derives from the public good. Individuals who are centrally located in the network thus exert more effort, which implies that aggregate effort and public goods provision will be increasing in the connectedness (density) of the network.

In our model, the level of effort exerted by each member of the coalition is fixed and all members receive the same benefit from participation (non-participants receive nothing). What is endogenous is the size of the largest stable coalition that can be supported in equilibrium, moving our analysis into the domain of the coalition formation literature (see Bloch, 2005, for a review on coalition size). We could, in principle, have modeled voting, the election of black leaders, and public goods provision during Reconstruction as a network game involving the entire population. However, this game would not describe the Great Migra-

[2] A number of measures of connectedness have been proposed in the networks literature. Closure (Coleman 1988) measures the overlap in the networks of connected agents. Clustering (Granovetter 1985) measures the extent to which two friends of a given agent are friends of each other. Support (Jackson et al. 2012) measures the number of pairs of friends that have some other friend in common. Finally, density (Karlan et al. 2009) is the ratio of the actual number of links to the potential number of links in the network. All of these measures will, in general, be positively associated with spatial proximity.

tion very well, where a group of individuals drawn from the network moved to the North, sharing the payoffs they generated amongst themselves but not with those that remained. An appealing feature of our model is that it is applicable to both Reconstruction and the Great Migration, while retaining its connection to the networks literature. Recall that the size of the largest stable coalition and, by extension, political participation and migration, is a specific nonlinear function of the connectedness in the social network from which it is drawn.

We derive and apply formal statistical tests of the theory relating network connectedness to the two outcomes of interest: black political participation during Reconstruction and the movement of black workers to northern cities during the Great Migration. Network connectedness is measured by black spatial proximity, which is constructed from postbellum cropping patterns. Although these cropping patterns, and their accompanying population distributions, can be traced back to antebellum decisions made by white landowners, blacks could have re-sorted after Emancipation. We verify that the results are robust to using that part of the variation in postbellum network connectedness that can be explained by (predetermined) antebellum cropping patterns.

Blacks could vote and elect their own leaders for a brief period during and just after Reconstruction (Morrison 1987, Foner 1988). They would have voted for the Republican Party (the party of the Union) at this time; and so black political participation in each county can be measured by the number of Republican votes. We find that Republican votes in national and state-wide elections during the 1870s match the specific nonlinearity, with a slope discontinuity, implied by the model. Since race-specific voting data are not available, we also examine the relationship between our crop-based measure of network connectedness and the probability that a black leader was elected by the county to the State Senate or House. The patterns match those of Republican votes, consistent with the presumption that black (Republican) voters would have wanted to elect members of their own race. Southern blacks were gradually disfranchised from the late 1880s through the 1890s as Jim Crow laws took effect. We find no association between network connectedness and Republican votes in 1900, which provides further evidence that the nonlinear voting patterns of the 1870s were primarily driven by blacks.

Although black disfranchisement was complete by 1900, a new opportunity to organize and work together arose with the Great Migration. Over 400,000 blacks moved to the North between 1916 and 1918 (exceeding the total number who moved in the preceding 40 years), and over one million left by 1930 (Marks 1989). The standard explanation for this movement, which varied substantially across southern counties, is that it was driven by the individual response to *external* factors that include the increased demand for labor in the wartime economy (Mandle 1978, Gottlieb 1987); the decline in cotton acreage due to the boll weevil invasion (Marks 1983); the segregation and racial violence that accompanied Jim Crow laws (Tolnay and Beck 1990); and the arrival of the railroads (Wright 1986, Black et al. 2012). Scant attention has been paid to the *internal* forces that would have supported network-based migration. This is surprising given the voluminous literature on networks in international and internal migration. Providing a new perspective on the Great Migration, we find that the relationship between network connectedness and various measures of black migration match the predictions of our network-based model once again - there is no correlation up to a threshold and a steep, monotonic association thereafter. This relationship is obtained during the period of the Great Migration but not before, providing us with a useful falsification test once again of the theory.

Our primary measure of out-migration is derived from changes in the black population in southern counties during the Great Migration, adjusting for natural changes due to births and deaths. We cross-validate this measure with another one constructed from newly-available data, which contain the city-of-residence in the 1970s (and after) of people born in Mississippi between 1905 and 1925, as well as the person's county of birth. While the year of migration is unknown, these data provide a direct measure of migration at the county level. This measure is highly correlated with the population change measure, and both variables show the same nonlinear association with network connectedness across Mississippi counties that we observe across all southern counties.

Since the Mississippi data contain the (final) destination city of each migrant, we can test another prediction of the theory. Migrants who are networked will move to the same place, whereas those who move independently will be spread across the available destinations. If variation in migration levels across southern counties is driven by differences in the size of underlying coalitions, then this implies that the *number* of black migrants and the spatial *concentration* of these migrants across destinations will track together. As predicted, the Herfindahl-Hirschman Index (HHI) of spatial concentration across destination cities for the Mississippi migrants is uncorrelated with network connectedness up to the same point as the level of migration, and steeply increasing in connectedness thereafter.

Having established a robust nonlinear relationship between our crop-based measure of network connectedness and both political participation and migration, the paper concludes by considering alternative explanations that do not rely on a role for networks. The first explanation posits that an external agency, such as the Republican Party or a northern labor recruiter, organized black voters or migrants. A related alternative explanation, which is most relevant during Reconstruction, is that blacks would only turn out to vote when they were sufficiently sure they could elect their own leader. As shown below, both explanations imply that voting and migration *levels* should shift discontinuously at a threshold, which is inconsistent with the data.

Another possibility is that individuals vote and migrate independently in response to external forces that vary across counties. A small white minority in counties where plantation crops were grown had enormous wealth during and after slavery. These are precisely the conditions under which the elite will co-opt political institutions to suppress wages and restrict mobility, and use violence and intimidation to achieve their objectives (Engerman and Sokoloff 1997, Alston and Ferrie 1999, Acemoglu and Robinson 2008). In response to these adverse conditions, blacks in plantation counties would have been more likely to vote during Reconstruction and to migrate to northern cities when job opportunities became available. If whites only organized above a threshold, then the observed nonlinear relationship between cropping patterns which we associate with black network connectedness and these outcomes could be explained as well. What this complementary mechanism cannot explain is the nonlinear relationship between the same cropping patterns and black church congregation size. This is our most direct measure of coalition size and is associated with cooperation within the black population. It also cannot explain the nonlinear association between the cropping patterns and the destination-city HHI of blacks from Mississippi. If blacks' migration decisions were based on factors that did not include a coordination externality, then the probability of moving to the same destination would not track migration levels so closely.

Our statistical tests consistently find that outcomes associated with black networks – Republican votes, election of black leaders, black church size, black migration – have patterns that match the restrictions of our theory. Variables that should not be associated with black networks – railroad density, Republican votes after Reconstruction, white church size, and white migration – do not. The implied magnitudes of the network effects, after controlling for standard determinants of migration, are large - for example, over half of the migrants to the North came from the third of southern blacks who lived in the densest counties, while less than fifteen percent came from the third in the least dense counties. Although anecdotal evidence suggests that networks linking southern communities to northern cities did emerge (Gottlieb 1987, Grossman 1989), we are the first to identify and quantify network effects in the Great Migration, an event of great interest across many disciplines. This paper concludes by discussing the significance of this finding for the subsequent evolution of black communities in northern cities.

2 Institutional Setting

This section begins by describing two major opportunities that presented themselves to African-Americans in the postbellum period: (i) the opportunity to vote and elect their own leaders during and just after Reconstruction, 1870-1890 and (ii) the opportunity to migrate to northern cities during the Great Migration, 1916-1930. We subsequently discuss the construction of the network connectedness measure, which would have determined the collective response to these opportunities across southern counties. This section concludes with an initial description of the relationship between this measure and both political participation and migration.

2.1 Reconstruction

Three amendments to the Constitution, passed in quick succession after the Civil War, gave political representation to African-Americans. The 13^{th} Amendment, passed in 1865, abolished slavery. The 14^{th} Amendment, passed in 1866, granted full rights of citizenship to African-Americans. And the 15^{th} Amendment, passed in 1869, gave them the right to vote. This opportunity coincided with the Reconstruction Act of 1867, which put the Confederate states under military (Federal) rule for the next decade. Blacks voted in large numbers for the Republican party during this period and elected their own leaders (King 2001 a, b). But Southern Democrats began to reassert themselves soon after Reconstruction had ended, and southern states began passing legislation from the early 1890s that effectively eliminated blacks from the electorate by 1900 (Du Bois 1908, Morrison 1987, Valelly 2004).

Although external organizations such as the Freedmen's Bureau and the Union League were active during Reconstruction, the major impetus for African-American political participation came from within (Stampp 1966, Foner 1988).[3] "In record time they organized, sponsored independent black leaders, and committed themselves to active participation ... It was now possible for blacks to not only field candidates for election but to influence the outcome of elections by voting" (Morrison 1987: 35). During Reconstruction, as many as 600 blacks sat in state legislatures throughout the South.[4] While this political success is impressive, what is even more impressive is the discipline and courage shown by black voters in continuing to vote Republican in large numbers and to elect their own leaders through the 1880s and even into the 1890s, after Federal troops had left the South (Kolchin 1993).

Where did the black leaders come from? The church was the center of community life in the postbellum period and it was natural that black political leaders would be connected to this institution (Du Bois 1908, Woodson 1921, Frazier 1964, Dvorak 1988, Valelly 2004). "... preachers came to play a central role in black politics during Reconstruction ... Even those preachers who lacked ambition for political position sometimes found it thrust upon them" (Foner 1988:93). African-American communities did not passively support these leaders. The

[3]At its peak in 1866, the Freedmen's Bureau employed only 20 agents in Alabama and 12 in Mississippi. It ceased most of its activities by the end of 1868 and was officially abolished in 1872, before black political participation even began (Kolchin 1993).

[4]About 50% of South Carolina's lower house, 42% of Louisiana's lower house, and 29% of Mississippi's lower house was black during Reconstruction. The corresponding statistics for the upper house were 19% in Louisiana and 15% in Mississippi. Blacks accounted for a sizeable fraction of state legislators even in states such as Virginia that did not witness a "radical" phase during Reconstruction (Valelly 2004).

political support they provided gave them benefits in return, until they were disfranchised towards the end of the nineteenth century (Morrison 1987).

2.2 The Great Migration

The first major movement of blacks out of the South after the Civil War commenced in 1916. Over the course of the Great Migration, running from 1916 to 1930, over one million blacks (one-tenth the black population of the United States) moved to northern cities (Marks 1983).[5] This movement was driven by both pull and push factors. The increased demand for labor in the wartime economy coupled with the closing of European immigration, gave blacks new labor market opportunities (Mandle 1978, Gottlieb 1987). Around the same time, the boll weevil invasion reduced the demand for labor in southern cotton-growing counties (Marks 1989). Adverse economic conditions in the South, together with segregation and racial violence, encouraged many blacks to leave (Tolnay and Beck 1990). Their movement was facilitated by the penetration of the railroad into the deep South (Wright 1986). A confluence of favorable and unfavorable circumstances thus set the stage for one of the largest internal migrations in history.

How did rural blacks hear about new opportunities in northern cities? The first links appear to have been established by recruiting agents acting on behalf of northern railroad and mining companies (Henri 1975, Grossman 1991). Independent recruiters, who charged migrants a fee for placing them in jobs, were soon operating throughout the South (Marks 1989). Apart from these direct connections, potential migrants also heard about jobs through ethnic newspapers. For example, the *Chicago Defender*, which has received much attention in the literature, increased its circulation from 33,000 in 1916 to 125,000 in 1918. Industries throughout the Midwest sought to attract black southerners through classified advertisements in that newspaper (Grossman 1991).

Although external sources of information such as newspapers and recruiting agents played an important role in jump-starting the migration process, and agencies such as the Urban League provided migrants with housing and job assistance at the destination, networks linking southern communities to specific northern cities, and to neighborhoods within those cities, soon emerged (Gottlieb 1987, Marks 1991, Carrington, Detragiache, and Vishwanath 1996). "[These] networks stimulated, facilitated, and helped shape the migration process at all stages from the dissemination of information through the black South to the settlement

[5]There were three phases in the Great Migration: an initial phase, 1916-1930; a slow down in the 1930s; and a subsequent acceleration, 1940-1970 (Carrington, Detragiache, and Vishwanath 1996). We focus on the initial phase, as do many historians (e.g. Mandle 1978, Gottlieb 1987, Marks 1989) because we are interested in black cooperation in the decades after Emancipation. Future work discussed in the concluding section will trace the evolution of these networks in northern cities over the course of the twentieth century, linking our project to previous contributions in urban economics (e.g. Cutler, Glaeser, and Vigdor 1999, Boustan 2010).

of black southerners in northern cities" (Grossman 1991: 67).

Reconstruction was more radical and persistent in the deep South (Kousser 1974, Kolchin 1993). During the Great Migration, the heaviest black out-migration occurred in an area that had been dominated by the plantation cotton economy. "Some counties were characterized by extremely high out-migration, while others maintained relatively stable black populations ... Such intra-state variation raises interesting questions about the causes of the differential migration ... Was the cotton economy particularly depressed? Were blacks subjected to more brutal treatment by whites in those areas? Did economic competition between blacks and whites restrict economic opportunity, and thereby encourage out-migration?" (Tolnay and Beck 1990: 350). Our novel explanation for (part of) this variation across counties is based on internal rather than external forces. Black population densities would have been larger in counties where a greater fraction of land was allocated to labor intensive plantation crops (not just cotton). The social interactions resulting from spatial proximity would have given rise to more connected networks, larger coalitions and, therefore, greater political participation during Reconstruction *and* larger population flows during the Great Migration.

2.3 Black Networks

Networks can only function effectively if their members interact with one another sufficiently frequently over long periods of time. Forced separation would naturally have made it difficult to support networks in the slave population (Du Bois 1908, Frazier 1939).[6] Nevertheless, the slave quarter and the independent informal church that often formed within the quarter, have been identified as domains within which cooperation, mutual assistance, and black solidarity did emerge (Blassingame 1972, Genovese 1974). "[Large plantations] permitted slaves to live together in close-knit communities – the slave quarters – where they could develop a life of their own" (Fogel 1989: 170). Most slaveholdings were too small to support such communities and interactions across plantations were relatively infrequent (Stampp 1956). Networks that covered a substantial area and linked a sizeable population could thus have only formed after Emancipation, once the restrictions on social interactions were lifted.

Our objective in the empirical analysis is to estimate the relationship between network connectedness and outcomes associated with black coalition formation. Network connectedness is determined by spatial proximity. The relevant population when constructing the connectedness measure would be the population from which coalitions during Reconstruction and the Great Migration were drawn. The postbellum South was largely rural, with

[6]The inter-state slave trade frequently separated families and plantation communities. For example, close to one million slaves moved to southwestern cotton states between 1790 and 1860 as production of that crop boomed (Fogel 1989, Kolchin 1993). Although Fogel and Engerman (1974) estimate that 84 percent of the slaves that moved west migrated with their owners, most other historians assign much greater weight to slave sales (Tadman, 1989, for instance, estimates that sales accounted for 70-80 percent of the slave movement).

agriculture the primary productive activity. The coalitions that formed would thus have consisted almost entirely of black agricultural workers (and tenants).[7] Our connectedness measures will be based on this population.

A distinctive feature of the antebellum South was the unequal size of slaveholdings and the uneven distribution of the slave population across counties (Stampp 1956). One-quarter of U.S. slaves resided in plantations with less than 10 slaves, one-half in plantations with 10-50 slaves, and the remaining in plantations with more than 50 slaves (Genovese 1974). This variation arose as a natural consequence of geographically determined cropping patterns and the organization of production under slavery (Wright 1978, 1986). Where labor intensive plantation crops such as cotton, tobacco, rice, and sugarcane could be grown, slaveholdings and the slave population density tended to be large. However, a substantial fraction of slaves lived in counties with widely dispersed family farms growing wheat or corn, where both whites and blacks worked on the land (Genovese 1974).[8]

Following the Civil War, while many blacks did move, most did not abandon their home plantations and those who did traveled only a few miles (Mandle 1978, Foner 1988, Steckel 2000).[9] The black population distribution remained stable, with the county-level population correlation between 1860 and 1890 as high as 0.85. We will also see that antebellum cropping patterns strongly predict postbellum patterns. This implies that black agricultural workers would have lived and worked in close proximity to each other in counties where plantation crops – cotton, tobacco, rice, and sugarcane – were grown historically and continued to be grown.

Our first connectedness measure is simply the fraction of land allocated to plantation crops.

$$\mathbf{M1}.\ S_i = \sum_j \frac{A_{ij}}{A_i},$$

where S_i measures spatial proximity in county i, A_{ij} is the acreage allocated to plantation crop j and A_i is total acreage. Although crop acreage at the county level is available from the 1880 census onward, our baseline connectedness measure is constructed in 1890, midway between Reconstruction and the Great Migration. We will verify that the results are robust to using the average of this measure over the 1880-1900 period.

[7] For example, most historical accounts of the Great Migration, with the exception of Marks (1989), suggest that the bulk of the migrants to northern cities were rural peasants.

[8] While just one or two slaves worked on a family farm growing wheat or corn, approximately 100 slaves worked on a rice or sugarcane plantation, 35 on a cotton plantation, and a somewhat smaller number on tobacco plantations (Fogel 1989).

[9] Federal assistance to former slaves who sought to acquire land was extremely limited (Kolchin 1993). 40,000 blacks in Georgia and South Carolina were granted land for homesteading by General Sherman in 1865, but the land was returned to their original owners by President Johnson. Similarly, only 4,000 blacks, most of whom resided in Florida, benefited from the Homestead Act of 1866. Apart from these limited opportunities, white landowners could also have actively discouraged black sharecroppers and laborers from moving (Naidu 2010).

Our second connectedness measure accounts for differences in labor intensity across plantation crops,

$$\textbf{M2}.\ \ S_i = \sum_j \beta_j \frac{A_{ij}}{A_i},$$

where β_j is the labor intensity (workers per acre) for plantation crop j obtained from farm studies conducted in the postbellum period. Notice that the M2 measure can be interpreted as the acreage weighted average of the labor intensity (spatial proximity) across all crops, with the implicit assumption that the black labor intensity on non-plantation crops is small enough to be ignored. To validate this assumption and to obtain independent estimates of the labor intensities, we estimate the following equation over the 1880-1900 period:

$$P_{it} = \sum_j \beta_j A_{ijt} + \alpha(A_{it} - \sum_j A_{ijt}) + f_i + \epsilon_{it}, \tag{1}$$

where P_{it} is total black population in county i in year t, f_i is a county fixed effect (to account for the population that is not connected to agriculture), and ϵ_{it} is a mean-zero disturbance term. One limitation of using population as the dependent variable is that the estimated coefficients reflect not just the number of workers on different types of land, but also their dependents. We will ignore this issue for the moment and continue to treat the estimated coefficients as labor intensities. The labor intensities obtained from previously published farm studies and the β estimates from the equation above are reported in Table 1, Columns 1-2. The α coefficient, which measures black labor intensity on non-plantation crops, is an order of magnitude smaller than the β coefficients. As assumed, blacks were largely engaged in the cultivation of plantation crops. Our third connectedness measure is the same as M2 except that we use the estimated β coefficients from equation (1),

$$\textbf{M3}.\ \ S_i = \sum_j \hat{\beta}_j \frac{A_{ij}}{A_i}.$$

Our fourth connectedness measure uses the complete count from the 1880 census, available from IPUMS-USA (Ruggles et al. 2010), which allows us to compute the number of black men, by age and occupation, residing in each southern county. Three occupations: (i) farm laborers and wage workers, (ii) farm owners and tenants, and (iii) unspecified laborers, account for 82 percent of black men aged 18-50. Individuals in the first two occupations will certainly belong to the network of agricultural workers (and tenants). Given the importance of agriculture in the postbellum southern economy, a large fraction of individuals in the third occupation would also have been directly or indirectly engaged in agriculture. We thus replace total black population 1880-1900, P_{it}, with the number of working-age black men in these occupations as the dependent variable in equation (1). Since we now have a single time period, the t subscripts and the fixed effects, f_i, are dropped when estimating the equation.

To be consistent with the previous connectedness measures, we continue to use 1890 acreage to measure A_{ij} and A_i.

The coefficients estimated with this equation, which can now be more appropriately interpreted as labor intensities, are reported in Table 1, Column 3. The β coefficients continue to be large and precisely estimated, while the α coefficient is once again an order of magnitude smaller than the β coefficients. Our final **M4** connectedness measure is constructed using the same specification as **M3** above, with the estimated β coefficients from 1880 rather than over the 1880-1900 period.[10]

The acreage allocated to different crops must satisfy the orthogonality condition for consistent estimates of the labor intensities to be obtained. Since this assumption cannot be validated, we use the M2 measure, where the labor intensities are based on farm studies, as the baseline measure in all the analysis that follows. We will, however, verify that the results are stable across all four measures. To facilitate this comparison, M2, M3, and M4 will be normalized to have the same mean and standard deviation as the fraction of land allocated to plantation crops, M1, the statistic that is most easy to interpret. This normalization, which simply involves multiplying each measure by a constant and then adding another constant term, has no effect on the shape of the relationship between each connectedness measure and the outcomes of interest. Nevertheless, it emphasizes the link between network connectedness and cropping patterns and we will often refer to our connectedness measure as the *plantation share* in the discussion that follows.

Figure 1 describes the baseline network connectedness measure (M2) in the 15 southern states in which slavery existed prior to Emancipation.[11] The message to take away from the figure is that there is substantial variation in this statistic across states and, more importantly, across counties within states. We will take advantage of this variation to include state fixed effects in all the results that we report, although the results are very similar with and without fixed effects. Figure 2A provides preliminary evidence on the relationship between plantation share and both political participation and migration. Political participation is measured by the number of Republican votes in the county in the 1872 presidential election, at which point in time blacks could freely vote and elect their own leaders. Migration is measured by black population change in the county from 1910 to 1930 minus the corresponding change from 1890 to 1910 (to control for natural changes in population across counties, as described below). The nonparametric regressions presented in Figure 2A reveal a highly

[10]To verify that individuals working in other occupations should not be included in the network of agricultural workers from which the coalitions were drawn, we replaced the number of working-age black men in the three occupations listed above with the number in all other occupations in 1880 as the dependent variable. The estimated coefficients are statistically and economically insignificant.

[11]The slave states are Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, South Carolina, Tennessee, Texas, and Virginia. Among these states, Kentucky, Missouri, Delaware, and Maryland did not join the Confederacy.

nonlinear relationship between plantation share and both outcomes.[12] Nonlinearities are commonly generated in models with network effects because there is an externality associated with individual participation. The model that we develop below will provide a simple explanation, based on differences in the size of underlying coalitions across counties, for the nonlinearity we have uncovered in Figure 2A. It will also generate additional predictions that we can take to the data.

3 Theory with a Test

The model developed in this section places additional structure on the nonlinear relationship observed in Figure 2A. We subsequently proceed to develop statistical tests of the model's predictions. These tests will be used in Section 4 to formally validate the model and to rule out alternative explanations for the empirical results that are obtained.

3.1 Individual Payoffs

There are many economic environments in which individuals cooperate to achieve a common objective. For example, a group of individuals could form a cooperative to work together and jointly produce a good. Alternatively, a group of individuals could form a mutual insurance arrangement, pooling their incomes to smooth consumption on the basis of a pre-specified sharing rule. In the applications that we consider, a group of black activists in a southern county could have come together to provide a service to a local political leader during Reconstruction. Members of the group would have canvassed potential voters and turned out themselves in local, state, and federal elections. Once the leader was elected, the coalition of activists would have worked on his behalf, helping to provide goods and services to the electorate and increasing his chances of reelection. In return for these services, the coalition would have received a transfer to be shared by its members. Alternatively, a group of black migrants could have moved together to a northern city during the Great Migration, helping each other find jobs and working diligently as a team once they were employed. In a production environment where worker ability and effort were unobserved by firms, such mutual support and diligence would have resulted in improved employment prospects and favorable wages for the members of the coalition.

As in.Munshi (2003) and Ballester et al. (2006), the payoff W received by each member

[12] Following Robinson (1988), state fixed effects are partialled out nonparametrically using a two-step procedure in Figure 2A and all the figures that follow. In the first step, the outcome under consideration (political participation or migration) and each state dummy is separately regressed nonparametrically on plantation share. The residual from the first regression is then regressed on the residuals from the state-dummy regressions. Using the estimated state coefficients, the state dummies are differenced from the outcome under consideration. This differenced variable is nonparametrically regressed on plantation share in the second step.

of the coalition is increasing in its size, N. We assume that the size effects are declining at the margin, perhaps due to congestion. Network connectedness, λ, is introduced in the model by assuming that it makes the coalition function more efficiently. Although we do not observe selection into the coalition, we expect more connected networks to give rise to more connected coalitions. Members of a connected coalition will work better together. A connected coalition will also support stronger collective punishments and, hence, larger *ex post* transfers that encourage members to help each other and generate superior payoffs.

For analytical convenience, let N and λ be real numbers. The payoff each individual receives from participation can then be expressed by the continuous function $W(N, \lambda)$. Based on the discussion above, the payoff function is increasing in N but at a decreasing rate; $W_N(N, \lambda) > 0$, $W_{NN}(N, \lambda) < 0$. If collective punishments are increasing in the *number* of social links, for example, then the efficiency gain from belonging to a connected coalition will be increasing in N. This implies that the cross-partial with respect to λ will be positive, $W_{\lambda N}(N, \lambda) > 0$. These (reasonable) restrictions on the payoff function will allow us generate results that are consistent with Figure 2A.

Let P be the population in the local area, which is defined to be small enough that only a single coalition can form. Individuals outside the coalition operate independently and we normalize so that their payoff is zero. Using the payoff in autarky as the benchmark, this implies the following limit condition:

C1. $\lim_{\lambda \to 0} W(N, \lambda) = 0 \ \ \forall N$

This is just saying that there is no additional payoff from belonging to the coalition, regardless of its size, when network connectedness becomes infinitesimally small ($\lambda \to 0$) and all members of the coalition are socially unconnected.

3.2 Maximum Stable Coalition Size

Given the payoffs described above, we now proceed to derive the maximum stable coalition size, N, that can be supported in a local area. Network connectedness, λ, varies exogenously across local areas, which are otherwise indistinguishable. Our objective is to derive the relationship between λ and N. During Reconstruction, N would refer to the number of activists who would have worked together to support the local political leader. During the Great Migration, N would refer to the number of individuals who moved as a group to the North. Although migration is a dynamic process, we can think of N as the stock of individuals who had moved and were providing mutual support to each other at a given point in time.

Since $W(N, \lambda)$ is increasing in N and we have normalized so that the payoff in autarky is zero, what prevents the entire population from joining the coalition? To place bounds on the size of the coalition, we assume that each member incurs a private effort cost c each period. Benefits are received up front by the coalition, with the expectation that each member will exert effort *ex post*. This could well describe the timing of wage setting and work effort in northern jobs, as well as the sequence of transfers (patronage) and community effort during Reconstruction. The commitment problem that arises here is that a self-interested individual will renege on his obligation in a one-shot game. This problem can be avoided if the coalition is active over multiple periods. Based on the standard solution to an infinitely repeated game, cooperation can be sustained if individuals are sufficiently patient, i.e. if the discount factor δ is large enough so that the following condition is satisfied:

$$\frac{W(N,\lambda) - c}{1-\delta} \geq W(N,\lambda).$$

The term on the left hand side is the present discounted value of cooperation for each member of the network. The right hand side describes the payoff from deviating. In the first period, the deviator receives the usual per capita payoff without incurring the effort cost. Although effort is not observed immediately, shirking is ultimately revealed at the end of the period. A single coalition operates in each county and the usual assumption is that deviators will be excluded from the group forever after. Since individuals operating independently receive a zero per-period payoff, the continuation payoff is set to zero.

Collecting terms, the preceding inequality can be written as,

$$W(N,\lambda) \geq \frac{c}{\delta}.$$

From condition C1, this inequality cannot be satisfied for $\lambda \to 0$ even if the entire population joins the network. This implies that all individuals must operate independently. As λ increases, there will be a threshold λ^* satisfying the condition,

$$W(P,\lambda^*) = \frac{c}{\delta}.$$

As λ increases above λ^*, this condition can be satisfied for smaller coalitions because $W_{\lambda N}(N,\lambda) > 0$. But we are interested in the *largest* stable coalition. It follows that the entire population will join the coalition for all $\lambda \geq \lambda^*$. This unrealistic result is obtained because the continuation payoff – set to zero – is independent of N. If cooperation can be sustained for a given coalition size N, it follows that it can be sustained for any coalition size larger than N. Thus, if cooperation can be sustained at all, the entire population will participate.

Genicot and Ray (2003) face the same problem in their analysis of mutual insurance. If individual incomes are independent, then a larger coalition does a better job of smoothing risk, and absent other constraints the entire population should join the insurance arrangement. Genicot and Ray consequently turn to an alternative solution concept, the coalition-proof Nash equilibrium of Bernheim, Peleg, and Whinston (1987), to place bounds on the size of the group and we will do the same. An appealing and more realistic feature of this Nash equilibrium refinement in the context of collective arrangements is that it allows sub-groups rather than individuals to deviate. The continuation payoff is no longer constant because deviating sub-groups can form arrangements of their own and we will see that this pins down the maximum size that the coalition can attain.[13]

The coalition-proof Nash equilibrium places two restrictions on deviating sub-groups: (i) only credible sub-groups, i.e. those that are stable in their own right, are permitted to pose a threat to the coalition. (ii) Only subsets of existing coalitions are permitted to deviate.[14] The condition for cooperation can now be described by the expression,

$$\frac{W(N,\lambda)-c}{1-\delta} \geq W(N,\lambda) + \frac{\delta}{1-\delta}\left[W(N',\lambda)-c\right],$$

where N' is the size of the deviating sub-group. Collecting terms, the preceding condition can be expressed as,

$$W(N,\lambda) - W(N',\lambda) \geq \frac{1-\delta}{\delta}c.$$

The greatest threat to a group will be from a sub-group that is almost as large, $N - N' \to 0$. For analytical convenience assume that c is an infinitesimal number.[15] If c is of the same order as $N - N'$, the ratio $\tilde{c} \equiv c/(N-N')$ will be a finite number. Dividing both sides of the preceding inequality by $N - N'$, the condition for cooperation is now obtained as,

$$W_N(N,\lambda) \geq \frac{1-\delta}{\delta}\tilde{c}.$$

For a given λ, the left hand side of the inequality is *decreasing* in N since $W_{NN}(N,\lambda) < 0$. This implies that there is a *maximum* coalition size above which cooperation cannot be sustained for each λ (if cooperation can be sustained at all as discussed below). This also ensures that the deviating sub-group of size N' will be stable, as required by our solution concept.[16]

[13] The canonical efficiency wage model solves the commitment problem by making the employer and the individual worker interact repeatedly and by allowing the wage to adjust so that the gain from shirking in any period is just offset by the loss in future (permanent) income. In our model, the size of the group and, hence, the per capita payoff adjusts so that participants are indifferent between working and shirking.

[14] Members of the deviating sub-group could, in principal, form a new coalition with individuals who were originally operating independently. Bernheim, Peleg, and Whinston justify the restriction they impose on the solution concept by arguing that asymmetric information about past deviations would prevent insiders and outsiders from joining together.

[15] This assumption, together with the assumption that N is a real number, allows us to differentiate the W function below. If we allowed c to be a finite number and N to be an integer, we would need to difference instead of differentiating, but it is straightforward to verify that the results that follow would be unchanged.

[16] Solve recursively to establish this result. Start with the smallest possible coalition of size N''. Deviators from this coalition are individuals and so are stable by definition. From the concavity of the $W(N,\lambda)$ function, the condition for cooperation will be satisfied for N'' if it holds for some $N > N''$. This establishes that N'' is stable. Next, consider a coalition just larger than N''. Using the same argument establish that it is stable. Continue solving in this way until N' is reached to establish that it is stable.

Genicot and Ray show that the set of stable insurance arrangements is bounded above once they allow for deviations by sub-groups. Our model generates stronger predictions that match Figure 2.

Proposition 1. *Coalitions will not form below a threshold level of connectedness, $\underline{\lambda}$. Above that threshold, the maximum stable coalition size, N^*, is increasing in network connectedness, λ.*

To prove the first part of the proposition, we take advantage of condition C1, which implies that $\lim_{\lambda\to 0} W_N(N,\lambda) = 0$. Cooperation cannot be supported and networks will not form for small λ. As λ increases, $W_{\lambda N}(N,\lambda) > 0$ implies that there will be a threshold $\underline{\lambda}$ at which cooperation can be supported, but only for groups of infinitesimal size ($N \to 0$). Above that threshold, since N^* is the largest group that can be supported in equilibrium for a given λ,

$$W_N(N^*,\lambda) = \frac{1-\delta}{\delta}\tilde{c}.$$

Applying the Implicit Function theorem,

$$\frac{dN^*}{d\lambda} = \frac{-W_{\lambda N}(N,\lambda)}{W_{NN}(N,\lambda)} > 0$$

to complete the proof.

Although Proposition 1 derives the relationship between λ and N^*, coalition size is not directly observed. We thus proceed to derive the relationship between λ and observed outcomes – political participation and migration – that are associated with underlying coalitions. Suppose that there are two types of individuals: type-1 individuals belong to, or are influenced by, the coalition, while type-2 individuals vote and migrate independently. The number of type-1 individuals is equal to, or weakly increasing in, $N^*(\lambda)$. If the number of type-2 individuals is independent of λ, then Proposition 1 can be restated as follows: political participation and migration will be *uncorrelated* with λ up to a threshold (not necessarily the same threshold) and *increasing* in λ thereafter. This is consistent with the patterns observed in Figure 2A, but places additional restrictions on the data which we test formally below.

We could assume, instead, that the number of type-2 individuals is increasing in the black population, which, in turn, is increasing in λ, since λ is measured by spatial proximity. The number of voters and migrants would then be increasing in λ below the threshold, falsely rejecting the theory. To account for this possibility, robustness tests will control for black population when estimating the relationship between λ and the outcomes of interest. We will see that the conditional estimates are very similar to the unconditional estimates, supporting the assumption that the number of type-2 individuals is independent of λ. We will maintain this assumption in the discussion that follows.

Multiple equilibria evidently exist above the threshold once we characterize individual participation in the coalition as the solution to a noncooperative game. Apart from the equilibrium derived above, no one participates in another equilibrium.[17] We assume in the analysis that follows that blacks were able to solve the coordination problem and so political participation and migration in each local area is based on the maximum stable size N^* derived above.

The restrictions we have placed on $W(N, \lambda)$ allow the model to generate results that are consistent with Figure 2A. Other assumptions or other models could generate different results, but this is not a concern as long as they do not match the figure. For example, we will see below that models in which blacks vote opportunistically above a threshold or in which an external agency organizes black voters and migrants can generate a *level* discontinuity. However, the *slope* discontinuity that is implied by our model turns out to be more difficult to obtain.

3.3 An Additional Implication of the Model

The model generates predictions for variation in the *level* of political participation and migration across local areas. It can be extended to generate predictions for the *distribution* of migrants across northern destinations. Let the number of type-2 migrants who move independently from each county be n. Assume that these migrants are distributed evenly across $M \geq 2$ destinations. The number of type-1 migrants who belong to the coalition will be zero below the threshold, $\underline{\lambda}$, and $N^*(\lambda)$ above the threshold. When the coalition does form, these individuals will move as a group to a single destination.

The Herfindahl-Hirschman Index, which is defined as the sum of the squared share of migrants across all destinations, can then be used to measure the concentration of migrants in the North. Below the threshold, the Herfindahl-Hirschman Index, $H(\lambda) = M\left[\frac{n/M}{n}\right]^2 = 1/M$, is uncorrelated with network connectedness, λ. Above the threshold,

$$H(\lambda) = \left[\frac{\frac{n}{M} + N^*(\lambda)}{n + N^*(\lambda)}\right]^2 + (M-1)\left[\frac{\frac{n}{M}}{n + N^*(\lambda)}\right]^2.$$

Differentiating this expression with respect to λ,

$$H_\lambda(\lambda) = \frac{2(M-1)\frac{n}{M}N^*(\lambda)N^*_\lambda(\lambda)}{[n + N^*(\lambda)]^3} > 0,$$

since $N^*_\lambda(\lambda) > 0$ for $\lambda \geq \underline{\lambda}$ from Proposition 1. The specific nonlinear relationship between the level of migration and network connectedness that we derived in Proposition 1

[17] There are no other equilibria in this noncooperative game. In particular, a coalition smaller than the largest stable coalition is not an equilibrium because any individual operating independently would want to deviate and join it, making everyone better off without affecting its stability.

should apply to the distribution of migrants at the destination as well. This result will hold as long as networked migrants cluster more at the destination than individuals who move independently.

Figure 2B describes migration to northern cities from counties in the state of Mississippi as a function of our network connectedness measure. These data are constructed by merging Medicare records with social security records, as described below, allowing migrants from each Mississippi county during the Great Migration to be linked to northern destination cities. Providing independent support for the relationship we uncovered in Figure 2A across all southern counties, there is no association between plantation share and the level of migration up to the same threshold as in that figure, after which a monotonic relationship begins. More importantly, the level of migration and the concentration of migrants in northern cities, measured by the Herfindahl-Hirschman Index (HHI), track very closely together in Figure 2B.

3.4 Testing the Model

The model indicates that network connectedness has no association with political participation and migration up to a threshold and a positive association thereafter. The location of the threshold is *a priori* unknown. To test the predictions of the model we follow standard practice, e.g. Hansen (1999), to estimate a series of piecewise linear regressions that allow for a slope change at different *assumed* thresholds. The pattern of coefficients that we estimate, with accompanying t-ratios, will locate our best estimate of the *true* threshold and formally test the specific nonlinearity implied by the model.

Ignoring the state fixed effects to simplify the discussion that follows, the piecewise linear regression that we estimate for each assumed threshold, S, is specified as

$$y_i = \beta_0 + \beta_1 S_i + \beta_2 D_i(S_i - S) + \beta_3 D_i + \epsilon_i \tag{2}$$

where y_i is political participation or migration in county i, S_i is our measure of network connectedness – the plantation share – in that county, D_i is a binary variable that takes the value one if $S_i \geq S$, and ϵ_i is a mean-zero disturbance term. β_1 is the baseline slope coefficient, β_2 is the slope change coefficient, and β_3 is the level change coefficient (measuring the level discontinuity at the threshold). We will estimate this regression for a large number of assumed shares, in increments of 0.0001, over the range $[0, 0.25]$.

The slope coefficients, β_1 and β_2, can be directly linked to the predictions of the model: $\beta_1 = 0$ and $\beta_2 > 0$ at the true threshold. To derive the pattern of t-ratios on β_1 and β_2 that we expect to obtain across the range of assumed thresholds when the data generating process is consistent with the model, we generated a data set that consists of two variables: the actual plantation share in our southern counties, S_i, and a hypothetical outcome, $\tilde{y}_i$, that is constructed to be consistent with the model, with the true threshold set at 0.09.[18] To verify that the data we have generated match the model, we nonparametrically regress $\tilde{y}_i$ on S_i in Figure 3A. All the nonparametric regressions in this paper are estimated with a narrow bandwidth. Despite the noise that we have added to the outcome, a slope change near the "true" threshold, 0.09, is clearly visible in the figure.

Having generated data that match the model, we next proceed to estimate equation (2) sequentially over a large number of assumed thresholds. The t-ratios for the two slope coefficients, β_1 and β_2, are reported in Figure 3B for each of these assumed thresholds. The t-ratio for the baseline slope coefficient remains close to zero for all assumed thresholds below the true threshold and starts to increase thereafter. The t-ratio for the slope change coefficient starts close to zero, then increases steadily reaching a maximum well above two where the assumed threshold coincides with the true threshold, and then declines thereafter.

To understand why the t-ratios follow this pattern, return to Figure 3A and consider the piecewise linear regression line that would be drawn for an assumed threshold to the left of the true threshold. The best fit to the data at that assumed threshold sets $\hat{\beta}_1 = \hat{\beta}_3 = 0$ and $\hat{\beta}_2 > 0$. This implies that the t-ratio on the baseline slope coefficient will be zero and the t-ratio on the slope-change coefficient will be positive. Now suppose we shifted the assumed threshold slightly to the right. It is evident that we would continue to have $\hat{\beta}_1 = \hat{\beta}_3 = 0$ since there is no change in the slope to the left of the assumed threshold, but $\hat{\beta}_2$ would increase and the regression line would do a better job of fitting the data to the right of the threshold. The t-ratio on the baseline slope coefficient would remain at zero, while the t-ratio on the slope-change coefficient would increase. This would continue as the assumed threshold shifted gradually to the right until it reached the true threshold.

Once the assumed threshold crosses to the right of the true threshold, the piecewise linear regression line that best fits the data will set $\hat{\beta}_1 > 0$. Although the magnitude of the baseline slope coefficient will increase as the assumed threshold shifts further to the right, the regression line will do an increasingly poor job of fitting the data to the left of the threshold. This implies that the t-ratio on the baseline slope coefficient is not necessarily monotonically increasing to the right of the true threshold, although it must be positive. In practice, this t-ratio will increase monotonically with both political participation and migration.

To derive the corresponding change in the t-ratio for the slope change coefficient, recall that the hypothetical outcome increases linearly to the right of the true threshold. Once the

[18] The value of the hypothetical outcome in each county is obtained by setting $\beta_0 = 670$, $\beta_1 = 0$, $\beta_2 = 7700$, $\beta_3 = 0$, and $S = 0.09$ in equation (2) and then adding a mean-zero noise term. These parameter values are derived from a piecewise linear regression of Republican votes in the 1872 presidential election on plantation share, with state fixed effects and the break at 0.09. We set the true threshold at 0.09 to be consistent with our best estimate of that threshold using the joint-test that will be discussed below. The variance of the mean-zero noise term in the simulation is set to match the variance of the residuals from this piece-wise linear regression.

level change coefficient is introduced, which must now be positive, $\hat{\beta}_3 > 0$, this implies that the regression line to the right of the assumed threshold will perfectly fit the data, except for the noise we have added to the outcome. This line maintains the same slope, and continues to precisely match the data, as the assumed threshold shifts further to the right. However, since the regression line to the left of the assumed threshold is growing steeper and is less precisely estimated as the assumed threshold shifts to the right, the *slope-change* coefficient and the t-ratio on that coefficient will unambiguously decline.

The preceding discussion and Figure 3B tell us what to expect when the data are consistent with the model. They also locate our best estimate of the true threshold. This will be the assumed threshold at which the t-ratio on the baseline coefficient starts to systematically increase *and* the t-ratio on the slope change coefficient reaches its maximum value.[19] Given the noise in the outcome measures, however, it is sometimes difficult to assess whether or not the t-ratios match the predictions of the model. This motivates a joint test of the model's predictions, based on the two slope coefficients, which also provides us with a single best estimate of the true threshold's location.[20]

Following standard practice, the composite null, which we test by estimating equation (2) at each assumed threshold, is set up to be inconsistent with the model:

$$H_0 : \beta_1 \geq |\epsilon h| \ \ and \ \ \beta_2 = 0,$$

where ϵ can be arbitrarily small and h is a scale parameter. β_1 is thus bounded away from zero under the null, while β_2 is set to zero. $\hat{\beta}_1$ will be mechanically further away from zero when the outcome variable has a larger mean or variance. To make the joint-test comparable across outcomes, we thus set h to be the standard deviation of the outcome under consideration multiplied by a constant.

Given the outcomes that we consider in this paper, the following data generating processes are feasible when the null is rejected:

$$H_1 : (i)\beta_1 = 0 \ \ and \ \ \beta_2 > 0$$
$$(ii)\beta_1 = 0 \ \ and \ \ \beta_2 = 0$$
$$(iii)\beta_1 \geq |\epsilon h| \ \ and \ \ \beta_2 > 0.$$

The first data generating process is consistent with our model. With the second data generating process, there is no relationship between the outcome under consideration and plantation share. With the third data generating process, the outcome is increasing monotonically in plantation share (with a slope change at a threshold).

[19]We could alternatively have plotted the baseline and slope change coefficients instead, over the range of assumed thresholds. The advantage of the t-ratios is that they allow us to test and compare the model across multiple outcomes.

[20]We are grateful to Yuya Sasaki for his help in deriving the test.

If we replaced the *and* statement under the null with an *or* statement, the parameter space would expand and only our model would be feasible if the null were rejected. However, this would introduce a new problem because any test statistic constructed to test the composite hypothesis would have different values under the null, depending on which component was relevant. We consequently retain the *and* statement, but we will see that our joint-test statistic nevertheless has the power to distinguish between the alternative data generating processes when the null is rejected.

The joint-test statistic is constructed as follows:

$$T(\beta) = \phi\left(\frac{\beta_1}{h}\right)\beta_2,$$

where ϕ is a symmetric and continuous function that reaches its maximum value at zero and the h parameter once again ensures that deviations in $\hat{\beta}_1$ away from zero are penalized consistently across outcomes. By the delta method,

$$\sqrt{n}\left(T(\hat{\beta}) - T(\beta)\right) \xrightarrow{d} N\left(0, DT(\beta)VDT(\beta)'\right)$$

where $V = \begin{bmatrix} V_{\beta_1} & V_{\beta_1\beta_2} \\ V_{\beta_1\beta_2} & V_{\beta_2} \end{bmatrix}$ and $DT(\beta) = \left[\frac{1}{h}\phi'\left(\frac{\beta_1}{h}\right)\beta_2 \quad \phi\left(\frac{\beta_1}{h}\right)\right]$.

$T(\beta) = 0$, under the null H_0 because $\beta_2 = 0$. Substituting the expressions for V and $DT(\beta)$, under the null

$$\sqrt{n}T(\hat{\beta}) \xrightarrow{d} N\left(0, \left[\phi\left(\frac{\beta_1}{h}\right)\right]^2 V_{\beta_2}\right).$$

Dividing by the standard deviation and then squaring,

$$\frac{n\left[T(\hat{\beta})\right]^2}{\left[\phi\left(\frac{\beta_1}{h}\right)\right]^2 V_{\beta_2}} \xrightarrow{d} \chi^2_1.$$

Under the null, β_1 has a range of values. We select the "least favorable" null, $\beta_1 = |\epsilon h|$, which minimizes the value of the preceding statistic. If we do reject the null, this implies that we would reject the null for any $\beta_1 \geq |\epsilon h|$. Following standard practice when implementing the Wald test, we replace V_{β_2} with $\hat{V}_{\beta_2}$. Substituting the expression for $T(\hat{\beta})$, we arrive at the statistic that is used for the joint test of the model,

$$n\frac{\left[\phi\left(\frac{\hat{\beta}_1}{h}\right)\right]^2}{\left[\phi(\epsilon)\right]^2}\frac{\hat{\beta}_2^2}{\hat{V}_{\beta_2}} \xrightarrow{d} \chi^2_1.$$

Because ϵ can be arbitrarily small, we set ϵ equal to zero (to be conservative) when computing the joint-test statistic. We will reject the null hypothesis if this test statistic exceeds the critical value for the chi-squared distribution with one degree of freedom.

If the data generating process is consistent with the model, $\hat{\beta}_1 = 0$ for all assumed thresholds to the left of the true threshold. However, $\hat{\beta}_2$ is increasing as we shift closer to the true threshold *and* is more precisely estimated. This implies that our joint-test statistic will be increasing in magnitude as the assumed threshold moves closer to the true threshold. After reaching its maximum value at the true threshold, where we are most likely to reject the null, the statistic will drop rapidly to zero if the ϕ function and the scale parameter, h, together place sufficient penalty on deviations in $\hat{\beta}_1$ away from zero. Recall that $\hat{\beta}_2$ is declining and less precisely estimated as the assumed threshold shifts further to the right of the true threshold, reinforcing this effect. In contrast, the (multiplicative) joint-test statistic that we have constructed will be zero when the data generating process is consistent with model (ii), and we will not reject the null hypothesis, since $\beta_2 = 0$. We will not reject the null under model (iii) either, if β_1 is sufficiently large and the ϕ function places sufficient penalty on deviations from zero. Our test statistic thus distinguishes our model from other data generating processes when the null is rejected.[21]

Figure 3C reports the joint-test statistic across the entire range of assumed thresholds, in increments of 0.0001, with our simulated data. We use the density of the standard normal distribution to characterize the ϕ function and set h equal to the standard deviation of the outcome under consideration, in the simulation exercise and in the analysis that follows.[22] The joint-test statistic is increasing in the assumed threshold in Figure 3C until it reaches its maximum value near the true threshold (0.09), declining steeply thereafter. The 95 percent critical value for the chi-squared distribution with one degree of freedom is 3.84, which implies that we can reject the null hypothesis at conventional levels for a range of assumed thresholds around the true threshold. We are nevertheless most likely to reject the null hypothesis where the joint-test statistic reaches its maximum value, and this will be our best estimate of the true threshold.

An alternative criterion to locate the true threshold, as suggested by Hansen (1999) is the assumed threshold at which the sum of squared residuals in the piecewise linear regression is minimized. Based on our description of Figure 3B, this is precisely the point at which the t-ratio on the baseline slope coefficient starts to increase (from zero) and the t-ratio on the slope change coefficient starts to decline. With no additional controls in the regression, Hansen's threshold and the threshold obtained from our joint test will also coincide. This

[21] In the empirical analysis that follows, we will consistently reject the null hypothesis for outcomes associated with black networks, where the data generating process based on the t-ratio test is consistent with model (i). In contrast, we will not reject the null for other outcomes where the t-ratio test indicates that the data generating process is consistent with model (ii) or model (iii).

[22] The results are robust to different values of h. For example, we experimented with half and three-quarters of the standard deviation without substantively changing the results. In practice, we set h to be slightly smaller for outcomes associated with black networks. This makes $\phi(\hat{\beta}_1/h)$ smaller and so it is more difficult to reject the null hypothesis with those outcomes.

is evident in Figure 3D, where Hansen's Likelihood Ratio (LR) statistic, which is simply a normalization of the sum of squared residuals, declines as the assumed threshold shifts to the right until it reaches its minimum value at the true threshold, as in Figure 3C, before starting to increase once again.[23]

The advantage of the joint test is that it is tied directly to the theory, locating the threshold at which the estimated slope coefficients – $\hat{\beta}_1$ and $\hat{\beta}_2$ – match most closely with the predictions of the model, i.e. $\beta_1 = 0$ and $\beta_2 > 0$. In contrast, Hansen's test searches for a slope change ($\beta_2 > 0$) without placing the additional restriction that the slope to the left of the threshold should be zero ($\beta_1 = 0$). His test has less power and, moreover, cannot distinguish between our model and a data generating process in which the outcome is a monotonically increasing nonlinear function of plantation share (model (iii) above). It does, however, independently locate that threshold when the data generating process is consistent with our model.

4 Empirical Analysis

This section begins by describing the relationship between our measure of network connectedness, the plantation share, and the black population. We subsequently estimate the relationship between plantation share and outcomes associated with black coalitions: political participation, church size, and migration. The estimated relationship matches the predictions of the model at specific points in time when we expect these coalitions to have been active, for black outcomes but not for white outcomes. The empirical analysis concludes by considering alternative models and establishing that they can match some but not all the patterns observed in the data.

4.1 Black Population

Variation in voting and migration across counties is generated in our model by internal forces that restrict the size of the largest stable coalition that can form. Suppose, instead, that networks are absent, but the relationship between plantation share and black population matches the patterns in Figure 2A; i.e. population is constant up to a threshold and increasing thereafter. If a fixed fraction of the black population votes and migrates, this would explain the patterns in Figure 2A without a role for networks.

[23] Hansen's LR statistic is $N\frac{SSR_S - SSR_{MIN}}{SSR_{MIN}}$, where SSR_S is the sum of squared residuals at assumed threshold S, SSR_{MIN} is the minimum value of this statistic across all thresholds, and N is the number of observations. Hansen derives his test without a mean shift variable (by setting β_3 to zero). Although the results with the mean shift variable are noisier in Figure 3D and with all the outcomes that follow, the relationship between the sum of squared residuals or, equivalently, the LR statistic and the assumed threshold is qualitatively the same.

To examine this possibility, we nonparametrically regress black population on plantation share in Figure 4A at four points in time: 1860, 1870, 1890, and 1910. It is apparent from the figure that black population is monotonically increasing in plantation share. The slope also gets steeper over time, perhaps due to higher fertility in the high plantation share counties or migration into those counties, and we will return to this observation when constructing the migration statistics below.

To formally verify that black population is monotonically increasing in plantation share, we report the results of the t-ratio test for 1870 (the other years are the same) in Figure 4B. If the relationship were linear, the t-ratio on the baseline slope coefficient would be greater than two and constant, while the slope change coefficient would be zero. What we see is that the baseline slope coefficient is always greater than two and that it is increasing in plantation share. The t-ratio on the slope change coefficient is also increasing in plantation share and just exceeds two before it starts to decline. While this indicates that the relationship is nonlinear, the important point is that black population is still increasing in plantation share everywhere. The joint test, reported in Figure 4C, confirms this conclusion. We cannot reject the null anywhere in each of those years, presumably because the ϕ function places sufficient penalty on deviations in $\hat{\beta}_1$ away from zero.

Our analysis takes advantage of the fact that cropping patterns in the South were initially put in place by white landowners. Variation in black spatial proximity in the antebellum period was an unintended consequence of those decisions. Once blacks were free, however, they could have moved to counties where networks were stronger, changing existing cropping patterns. To account for such endogenous sorting as well as for differential fertility across counties in the postbellum period, we estimate regressions that use that part of the variation in 1890 plantation share that can be explained by 1860 cropping patterns.

The two-step estimation procedure that we implement is based on the nonparametric instrumental variable procedure suggested by Newey, Powell, and Vella (1999), except that we do not claim that 1860 cropping patterns satisfy the exclusion restriction. In the first step, we regress 1890 plantation share on a full set of state dummies and a quartic function of the median slaveholding in 1860.[24] Although crop acreage at the county level is not available prior to 1880, crop production data go back to 1840. In an alternative first-stage specification, we regress 1890 plantation share on a cubic function of 1860 production for each of the four plantation crops, additively and without interaction terms. The goodness of fit (R-squared) in the first-stage regressions exceeds 0.5, consistent with the view that

[24]The 1860 population census provides the number of slaveholdings by size-category in each county. These categories are all integers up to 9, 10-14, 15-19, 20-29, 30-39, 40-49, 50-69, 70-99, 100-200, 200-300, 300-500, and greater than 500. A hypothetical ranking of all slaves in a county can be constructed based on the size of the plantation to which they were assigned, which allows us, in turn, to compute the size of the slaveholding associated with the median slave.

postbellum cropping patterns could be traced back to antebellum decisions. In the second step, we nonparametrically regress 1870 black population on plantation share, partialling out state fixed effects (as usual) as well as a flexible (fifth-order polynomial) function of the first-stage residuals. By including the first-stage residuals in the second step, we effectively regress black population on the predicted plantation share (based on 1860 cropping patterns) and we see in Figure 4D that the relationship is more linear than what we obtained in Figure 4A with both instruments. This could be because the two-step procedure corrects for postbellum sorting across counties or because it purges measurement error.

Black population is increasing monotonically in plantation share. The increase in voting and migration above a plantation share threshold that we will observe below thus can be explained by the accompanying increase in the black population. The absence of such an association below the threshold, despite the fact that the black population is getting larger, is less easy to explain without a role for cooperation.

4.2 Reconstruction

A larger coalition of black activists in a county during Reconstruction would have generated greater political participation in the population. While political participation by race is not available, voter turnout by party is available from U.S. Historical Election Returns (ICPSR). Because blacks would have voted Republican at this time, our primary measure of political participation is the number of Republican votes in the county. This statistic is reported at three points in time in Figure 5A, for the 1872, 1880, and 1900 presidential elections. The pattern of votes in 1872, which is at the height of black political power, was reported earlier in Figure 2. Although Southern Democrats started to take control and blacks were gradually disfranchised once Reconstruction ended in 1877, blacks continued to vote and to elect their own leaders, with less and less success, into the 1890s. As expected, the increase in Republican votes past the plantation share threshold is weaker in 1880 than in 1872. However, the specific nonlinearity implied by the model continues to be obtained. This contrasts with the pattern in 1900, by which point in time blacks would have been completely disfranchised and where we see no relationship between the number of Republican votes and plantation share.

Figure 5B tests whether the nonlinear relationship that we uncovered in 1872 in Figure 5A matches the model. The t-ratio on the baseline slope coefficient is close to zero up to a threshold plantation share and increasing thereafter. The t-ratio on the slope change coefficient increases steadily up to the same threshold, reaching a maximum value of four, and then declines thereafter. Figure 5C reports the joint-test statistic across the range of plantation shares in 1872 and 1900. The 1872 statistic reaches its maximum value, well above the 95 percent critical value for the chi-squared distribution with one degree of freedom, close to the threshold in Figure 5B. It declines steeply, on both sides, away from our best estimate of the true threshold (around 0.09). The 1900 statistic, in contrast, is close to zero across the entire range of assumed thresholds. Figure 5D reports the nonparametric instrumental variable estimates for 1872 Republican votes. The results are even stronger (and the relationship between voting and plantation share is more nonlinear) than what we obtained in Figure 5A. The results with political participation match the model's predictions and the simulations in Figures 3B and 3C, contrasting with what we observed for black population.

We next proceed to establish the robustness of these results to alternative measures of network connectedness and black voting. Figure 5E reports nonparametric estimates where the M2 measure is replaced by (i) the average of that measure over the 1880-1900 period, and (ii) the M3 and M4 measures. The estimated relationship is robust to the plantation share measure and the same result is obtained with other outcomes discussed below (not reported). Federal, state, and local elections are synchronized in the American political system and so the voter turnout across counties that we observe for presidential elections should also apply to local elections occurring at the same time, where the implications of the model may be more relevant. Figure 5F regresses Republican votes in gubernatorial and congressional elections (separately) on plantation share, uncovering the same pattern that we obtained with 1872 presidential elections.[25] The relationship between plantation share and Republican votes is robust to the type of election and we expect that the same relationship would be obtained with state and local elections, although those data are unavailable.

While the robust nonlinear relationship between Republican votes and plantation share we have uncovered is consistent with the model, we do not have direct evidence that the increase in Republican votes above the threshold was driven by black voters. White "carpetbaggers" from the North and white "scalawags" from the South also voted Republican in southern counties at this time. If the number of white Republican votes was correlated with plantation share, this could confound our interpretation of the results in Figure 5A. One observation from that figure that goes against this alternative explanation is that the number of Republican votes and plantation share are unrelated in 1900, by which time blacks were effectively disfranchised. To provide further support for our hypothesis, we take advantage of the fact that an increase in black votes would have generated an increase in black leaders, to the extent that blacks wanted to elect members of their own race.

Foner (1993) provides a complete list of black officeholders during Reconstruction. Almost all of these officeholders were elected to positions in state government. We therefore

[25]Republican votes in gubernatorial and congressional elections are available, by county, from ICPSR. Gubernatorial elections were held at four-year intervals but were not synchronized across states. Figure 5E is thus based on all gubernatorial elections held between 1871 and 1873. Data on congressional elections are obtained for 1872.

construct two measures of leadership based on his data: whether a black state representative and whether a black state senator was elected from each county in this period. These measures are regressed nonparametrically on plantation share in Figure 6A. The probability that a black leader, especially a state representative, was elected from a county tracks closely with the pattern of Republican votes in 1872 and 1880, indicating that voting patterns in those years were indeed being driven by black voters. Figure 6B reports the joint test statistic for state representatives (who accounted for most black leaders) and for Republican votes in gubernatorial elections (as a benchmark). Matching Figure 5C, which reports the corresponding statistic for the 1872 Presidential election, the test statistic reaches its maximum value (well above the critical value needed to reject the null with 95 percent confidence) at the same assumed threshold, just to the left of 0.1, for both outcomes. The same results are obtained for state senators (not reported).

The unconditional relationship between plantation share and political outcomes can only be used to test the theory, based on unobserved coalition size, if the number of individuals voting independently is uncorrelated with plantation share. This will not be the case if this number is increasing in the black population since we know from Figure 3 that black population is increasing in plantation share. To allow for this possibility, we estimate the *conditional* relationship between plantation share and the probability that a black state representative is elected in the county in Figure 6C. The black population in the county, together with the state fixed effects, are partialled out nonparametrically, generating results that match closely with the unconditional estimates.

Our explanation for the variation in voting across counties is based on internal forces that generate differences in the size of the largest stable coalition that can form (and encourage voter turnout). A related explanation is based on competition between blacks and whites. Consider a model of political competition in which blacks only turn out when they expect to win and elect their own leader with sufficiently high probability. Because black population and the share of blacks in the population (not reported) are both increasing in plantation share, blacks will not turn out to vote until a threshold share, which is consistent with the voting patterns that we observe. This model implies that there will be a discrete jump in voter turnout (sufficient to win the election) at the threshold. Formal tests reported below reject a mean shift at the threshold. Nevertheless, as a final robustness test, we condition nonparametrically for the black population share when estimating the relationship between plantation share and black leadership in Figure 6C.[26] Once again, the conditional results match closely with the unconditional estimates, indicating that our results are not being

[26]Black population and black population share are included (separately) as linear terms, and then partialled out nonparametrically, in Figure 6C. Similar conditional estimates (not reported) are obtained with a more flexible first-stage specification.

driven by strategic voting considerations.

4.3 The Black Church

A county can cover a large area. Given the high transportation and communication costs at the time, it would have been difficult for black residents across an entire county to work together. Coordination within *local* communities, with accompanying local political participation, would have aggregated up to the political participation we observe at the county level. As described in Section 2, community life in the postbellum period centered on the church, and African-American churches played an important political role during Reconstruction (DuBois 1908, Frazier 1964, Dvorak 1988). African-American politicians were often drawn from the clergy and church congregations worked together to support local leaders (Woodson 1921, Foner 1988). The level of local cooperation, characterized by the largest stable coalition in the model, can thus be conveniently measured by black church congregation size. This allows us to test the theory more directly at the local level and also provides micro-foundations for county-level political participation.

While slaves worshipped in biracial churches for the most part, they did appear to have some autonomy in the choice of denomination and most were formally affiliated with either the Baptist or Methodist church (Woodson 1921, Genovese 1974). Once free, they quickly formed independent congregations within those denominations (Boles 1988, Kolchin 1993). Southern blacks could remain part of the mainstream Baptist and Methodist denominations they belonged to as slaves, or they could affiliate with exclusively black sub-denominations, that spread throughout the South after the Civil War. Some of these sub-denominations, such as the African Methodist Episcopalian (AME) Church and the African Methodist Episcopalian Zion (AMEZ) Church, were established by freed blacks in northern cities at the beginning of the nineteenth century (Du Bois 1908). Black Baptist sub-denominations coalesced much later (Frazier 1964).

The Census of Religious Bodies (CRB) provides information on the number of churches in each county, by denomination, at roughly ten-year intervals from 1860 onwards.[27] We measure average congregation size in each denomination by the ratio of church members to the number of churches. The 1890 census is the only round in the postbellum period that recorded information on the number of church members in each denomination and that separately identified the black sub-denominations within the Baptists and Methodists. The advantage of having information on these sub-denominations is that the average congregation-size we compute for them will be based entirely on black congregations. Southern whites,

[27]The CRB was conducted as part of the population census from 1860 to 1890, with census enumerators collecting information from individual churches in each county. Subsequently, the U.S. Bureau of the Census conducted the CRB separately from the population census in ten-year intervals from 1906 to 1936.

like southern blacks, were most often Baptist or Methodist (Kolchin 1993). The average congregation-size that we compute for the Baptists and the Methodists as a whole will thus be based on black as well as white congregations. For this reason, the analysis of congregation-size that follows will be restricted to the 1890 census and will separately consider Baptists and Methodists, black sub-denominations among the Baptists and Methodists, and other (white) denominations such as the Presbyterians, Episcopalians, and Catholics.[28]

Figure 7A nonparameterically regresses average congregation size in each set of denominations described above on plantation share. The pattern for the Baptists and Methodists and for the black sub-denominations matches the corresponding pattern for black political participation and leadership that we obtained earlier: there is no association between average congregation size and plantation share up to a threshold and a positive association thereafter. Notice that the increase in congregation size past the threshold is greater for the black sub-denominations than for Baptists and Methodists as a whole. This implies that the results are not being driven by variation in the size of white congregations across counties. Consistent with this interpretation, no particular relationship between congregation size and plantation share is observed for other (white) denominations.

Figure 7B formally tests whether the nonlinear pattern observed in Figure 7A for the black sub-denominations is consistent with the model. The joint-test statistic increases steeply in the assumed threshold until it reaches its maximum value and declines steeply thereafter for both the black sub-denominations and for Baptists and Methodists. The maximum value of the joint-test statistic for Baptists and Methodists overall is well above the critical value needed to reject the null with 95 percent confidence, whereas it is close to zero over the entire range of assumed thresholds for the other denominations. Although we cannot reject the null hypothesis at conventional levels (we can with 90 percent confidence) with the black sub-denominations, our best estimate of the true threshold is close to what we obtained earlier for voting and black leadership. As with political participation, the results are stronger (and the relationship between black church congregation size and plantation share is more nonlinear) when we instrument for 1890 plantation share with 1860 median slaveholding and 1860 plantation-crop production in Figure 7C. The analysis of church size thus provides micro-foundations, based on our most direct measure of coalition size, for variation in political participation and black leadership across counties.

[28]The black sub-denominations included in the 1890 CRB are Regular Baptist (colored), African Methodist Episcopal, African Methodist Episcopalian Zion, Colored Methodist Episcopalian, and Colored Cumberland Presbyterian. Among these sub-denominations, only the Cumberland Presbyterians, who had a small following, fell outside the umbrella of the Baptists and the Methodists.

4.4 The Great Migration

We next proceed to examine the relationship between plantation share and the level of migration across southern counties. Since the population census does not provide the county of birth, the birth-location of blacks residing in northern cities in 1920 and 1930 cannot be used to measure the level of migration from each southern county. The census survivor ratio method has been proposed in the historical demography literature to deal with this problem (e.g. Lee et al. 1957, Collins 1997). In our application, this method would predict what a southern county's population would have been at the end of a given decade in the absence of migration – based on the age and sex distribution, each cohort's survival ratio (determined by mortality at the national or regional level) and fertility (for the youngest cohort). The difference between this predicted population and the actual population would provide an estimate of intercensal migration. We do not use this procedure for two reasons. First, the age distribution is not available at the county-level for census rounds between 1870 and 1930 (except for the full count in 1880). Second, even if these data were available, a single survival ratio and fertility rate could not be applied since we saw in Figure 4A that the black population was growing more rapidly in high plantation share counties in the postbellum period.

Our approach – which we will validate with an independent migration measure discussed below – uses county-level population changes just prior to the Great Migration to "non-parametrically" predict the changes that would have occurred in the absence of northern migration during that period. The first major movement to the North commenced in 1916. The population change in the preceding decade, $P_{1900} - P_{1910}$, predicts the change that would have occurred in the next decade in the absence of migration. The "short" double-difference, $(P_{1910} - P_{1920}) - (P_{1900} - P_{1910})$, is thus our best estimate of northern migration in each county between 1916 and 1920. The "long" double-difference, $(P_{1910} - P_{1930}) - (P_{1890} - P_{1910})$ provides an analogous measure over the course of the Great Migration.

Figure 8A nonparametrically regresses the change in population, $P_{1910} - P_{1920}$ and $P_{1900} - P_{1910}$, separately for black and whites, on plantation share. $P_{1900} - P_{1910}$ for blacks is negative everywhere and mildly declining in plantation share. This implies that the black population was increasing on net throughout the South prior to the Great Migration, particularly in counties with large plantation shares, which is consistent with the changes over time observed in Figure 4A. This relationship is reversed in the subsequent decade. There is no population change up to a threshold plantation share and a large *decline* in the population thereafter, which we attribute to migration. In contrast, population change for the whites is stable over the two decades, providing a useful benchmark for the results we obtain for the blacks.

Figure 8B adjusts for natural population change by nonparametrically regressing the short double-difference, $(P_{1910} - P_{1920}) - (P_{1900} - P_{1910})$ on plantation share. Figure 8C repeats the exercise with the long double-difference, $(P_{1910} - P_{1930}) - (P_{1890} - P_{1910})$, as the dependent variable. The regression with the long double-difference was reported earlier in Figure 2A and we see that the same pattern is obtained with the short double-difference. There is no association between plantation share and our measure of black migration up to a threshold and a positive association thereafter. This contrasts with white migration, where a monotonic and mildly declining relationship with plantation share is observed. Figures 8D and 8E formally test the predictions of the model. The joint test-statistic with the short-double difference measure of black migration reaches its maximum value to the right of 0.1 in Figure 8D, although a second peak to the left of 0.1 is also visible. The corresponding statistic for the long double-difference measure reaches its maximum value to the left of 0.1 in Figure 8E, at a point that coincides with our best estimate of the threshold for voting and black church congregation size. The maximum value of the joint-test statistic is well above the critical value at which we can reject the null with 95 percent confidence in both figures. This contrasts with the test statistics for black population-change prior to the Great Migration as well as for white migration, which are close to zero across the range of assumed thresholds in both figures. As with voting and black church congregation size, the results are even stronger (and the relationship between black migration and plantation share more nonlinear) when we instrument nonparametrically for 1890 plantation share with 1860 slaveholding in Figures 8F and 8G. We also see in those figures that the estimates are robust to conditioning for variation in black population across counties.

Although it does account for natural population change, the double-differenced statistic is still an indirect measure of migration. To verify the robustness of the results in Figure 8 we consequently utilize newly available data from the state of Mississippi that link southern counties to northern destinations. These data include the zip code of residence of all recipients of Medicare Part B between 1976 and 2001. The Medicare records, which are reliably available from the 1905 birth-cohort onward, were merged with social security records (the Numident file), which include the town of birth. Under the assumption that individuals remained in the city (MSA) to which they moved, we can compute the number of migrants and the distribution of migrants across northern cities, by race, for each Mississippi county. These statistics are computed for individuals born between 1905 and 1925 because these are the individuals most likely to have migrated between 1910 and 1930, either as young adults or as children with their parents. While the large number of cohorts allows us to measure migration from each southern county with precision, this also implies that some individuals who moved after the Great Migration will be included in these cohorts. This will not qualitatively change the analysis that follows, because southern counties that channeled their members to particular northern destinations during the Great Migration would have continued to do so thereafter once communities were established.

Figure 9A nonparametrically regresses the short and long double-difference statistics that we use to indirectly measure migration, and a direct measure based on the 1905-1925 birth cohorts, on plantation share across Mississippi counties.[29] Reassuringly, these measures of migration track closely together and, moreover, match the pattern that was obtained across all southern counties. Although not reported, this pattern is obtained across Mississippi counties for Republican votes in 1872, the probability that a state representative was elected, and black church congregation size. Figure 9B reports nonparametric regressions with the number of migrants and the distribution of migrants, measured by the Herfindahl-Hirschman Index. As observed in Figure 2B, both statistics for blacks are uncorrelated with plantation share up to the same threshold and increasing in plantation share thereafter. In contrast, the number and the distribution of white migrants is uncorrelated with plantation share. The specific nonlinearity we have uncovered appears consistently across multiple outcomes associated with black coalitions at specific points in time. Notice that it is not obtained with other outcomes such as Republican votes after Reconstruction, church size in white denominations, white migration, and measures of black population change constructed prior to the Great Migration.

Tables 2 and 3 summarize the results we have obtained for outcomes associated with black networks. Table 2 reports regression estimates at the threshold obtained from the joint test, while Table 3 reports the estimates at the corresponding threshold from the Hansen test (with and without a mean shift at the threshold). The threshold locations from the joint test and the Hansen test match very closely. The baseline slope and the threshold mean shift coefficients are statistically indistinguishable from zero across all outcomes in both tables. In contrast, the slope change coefficient is positive and significant in both tables.

4.5 Alternative Explanations

The first alternative explanation for the results we obtain assumes that an external agency solves the commitment problem and organizes political participation during Reconstruction and the movement north during the Great Migration. Depending on the context, this agency could be the Republican party or a northern labor recruiter. The value to the agency $V(N)$ is an increasing function of the number of individuals, N, that it can mobilize. It is reasonable to assume that N is an increasing function of the black population of the county, which was shown to be increasing in the plantation share S. N is thus an increasing function of S, $N(S)$.

[29] All the nonparametric regressions up to this point in the analysis have included state fixed effects. Since we are now focussing on a single state, the two-step procedure used to partial out the state fixed effects is no longer required.

The alternative explanation can explain the increase in Republican votes and migration to the right of a plantation share threshold, simply because there is a larger black population to draw from. To explain the absence of such a relationship to the left of the threshold, introduce a fixed cost k. The external agency will only enter counties where it expects to mobilize a sufficiently large number of individuals. Because V is increasing in N, and N is increasing in S, there exists a threshold $\underline{S}$ below which there is no entry.[30] N is constant (zero) to the left of $\underline{S}$ and increasing in S to the right of $\underline{S}$.

This alternative *centralized* explanation has many features in common with our model of *decentralized* coalition formation. What distinguishes the alternative explanation from our model is a level discontinuity at the threshold (a discrete jump to $N(\underline{S})$) which is needed to just offset the fixed cost and which is not implied by our model. We do not observe a discrete jump at the threshold in any of the figures presented in this paper. What we observe instead is a change in the slope at the threshold. Formal tests of the model at our best estimate of the true threshold, reported in Tables 2 and 3, are also consistent with this observation.

The second alternative explanation that we consider assumes that individuals vote and migrate independently in response to external forces that vary across counties. For example, three push factors that have featured prominently in the literature on the Great Migration are the the arrival of the railroad, racial intimidation and violence, and the boll weevil invasion in cotton-growing counties. A well documented feature of the Great Migration is positive selection on education (eg. Lieberson 1978, Margo 1990, Tolnay 1998).

It is entirely possible that the strength of these push factors and other factors such as education that determined the response to new opportunities in the postbellum period varied with plantation share. Wealthy white landowners in the counties where plantation crops were grown would have benefited disproportionately by suppressing wages and restricting labor mobility (Engerman and Sokoloff 1997, Alston and Ferrie 1999, Acemoglu and Robinson 2008). One way to achieve this objective would have been through intimidation and racial violence (Tolnay and Beck 1990). A second strategy would have been to reduce public expenditures on black education in those counties (Margo 1990). As in our model of black coalition formation, suppose that white coalitions only form above a threshold plantation share, with their size (and, hence, their ability to exploit the black population) increasing in plantation share above the threshold. Then this would explain the results that we obtain, as the black response to white oppression when the opportunities for such a response became available during Reconstruction and the Great Migration.

We do not find evidence supporting this alternative explanation. Data on the number of black lynchings in each southern county between 1882 and 1915 (just before the onset of

[30]This threshold must satisfy the condition $V(N(\underline{S})) = k$. $V(N(S)) < k$ for $S < \underline{S}$ since $V_S(N(S)) > 0$ and so there is no entry below the threshold.

the Great Migration) indicate that very few lynchings are actually reported in this period.[31] Although the data are quite noisy, no relationship between plantation share and lynchings is detectable. Black literacy is declining mildly in plantation share, while the pattern for white literacy is reversed. However, the specific nonlinearity associated with white coalition formation, with a reduction in public expenditures on black education below a threshold, is not observed. Other push factors that have featured prominently in the literature, such as access to railroads and the effect of the boll weevil invasion, are also found to be uncorrelated with plantation share.[32] Recall that the statistic we use to test the model will be zero when the outcome under consideration is uncorrelated with plantation share. Consistent with the preceding discussion, the joint test statistic for all of these push factors, as well as for education, stays close to zero across the range of assumed thresholds in Figure 10.

While the push factors we consider may have been important for individual decisions to migrate, they cannot explain the specific nonlinear relationship between plantation share and black migration. It is possible that our measures do not fully capture the forces that were relevant during the Great Migration. For example, agricultural wages could have been declining steeply past a plantation share threshold as discussed above.[33] It is also possible that white landowners selected particular types of slaves in those counties. Any alternative explanation based on independent individual actions or characteristics would need to explain political outcomes many decades earlier. All of the push factors that we consider, for example, would not apply to political participation during Reconstruction.[34] More importantly, external forces that increased the propensity of blacks to migrate (independently) in some counties would not necessarily channel them to a restricted number of northern destinations. The observation that the level of migration and the concentration of migrants across destinations track closely together is difficult to explain without a model of underlying cooperation. Variation in the size of the black church congregation is also difficult to explain without a role for local coordination.

[31]These data are obtained from the Historical American Lynching (HAL) Data Collection Project. They do not contain observations from Delaware, Maryland, Missouri, Texas, and Virginia.

[32]Railroad access is measured by the number of miles of railroad in 1911 divided by the area of the county (available in 1880). The boll weevil invasion commenced in the cotton south around 1890, so we measure the boll weevil effect by the percentage change in cotton acreage from 1890 to 1920, at the onset of the Great Migration. Alternative measures, based on the percentage change from 1910 to 1920, as well as the percentage change in cotton production, generate similar results.

[33]It has been argued that the loss of economies of scale in agricultural production after slavery resulted in a dramatic decline in productivity (Fogel and Engerman 1974, Goldin and Lewis 1975, Moen 1992, Irwin 1994). This decline may have been particularly severe in the high plantation share counties. But this would not explain the high migration from those counties if productivity was equalized across counties in the postbellum period. What we would need is that wages were systematically suppressed in high plantation-share counties.

[34]The boll weevil invasion and the arrival of the railroad occurred after Reconstruction. Although blacks were quick to invest in education after Emancipation, slaves were largely illiterate (Du Bois 1908). We would thus expect little variation in black (adult) literacy rates across southern counties in 1872. Because the South was under Federal rule at that time, we would expect racial violence and intimidation to have been less relevant as well.

5 Conclusion

The development process has historically been characterized, and continues to be characterized, by the movement of groups across space and occupations. The analysis in this paper highlights the interaction between historical preconditions and new opportunities in shaping such group mobility. Despite the adverse circumstances that they faced under slavery, blacks were able to solve the coordination problem and respond as a group to new political and economic opportunities in the postbellum period. It is worth emphasizing, however, that the collective response we uncover is restricted to southern counties where specific preconditions, determined by the organization of agricultural production under slavery and thereafter, were satisfied. Over 50 percent of southern counties and one-third of the black population were situated below the threshold at which coalitions could form (at a plantation share around 0.09). These counties accounted for less than 15 percent of the blacks who migrated to northern cities.

Black migrants from counties below the threshold would have moved with relatively little social support. Blacks from counties above the threshold would have moved in large groups to a small number of northern destinations. This variation in the pattern of out-migration would have had consequences for the formation and evolution of black communities in northern cities. Relatively weak communities would have formed in destinations that received migrants who moved independently from diverse origin locations. In contrast, the small number of northern destinations that received the bulk of their migrants from southern counties above the threshold would have formed more cohesive communities. This variation in initial conditions would, in turn, have shaped the evolution of African-American communities over the course of the twentieth century.

Differential out-migration could also have had consequences for the evolution of black communities in southern counties. Given the well documented positive selection on education among northern migrants, counties above the threshold would have lost the bulk of their most able residents over the first half of the twentieth century. The resulting social dislocation could then explain Putnam's observation that those counties have relatively low social capital today. Wilson (1987) famously argued that the exit of educated black professionals from northern neighborhoods after Civil Rights and desegregation resulted in social dislocation and the concentration of poverty in inner-cities. A similar dynamic process may well have occurred in certain southern counties at the beginning of the twentieth century, paradoxically because they were better positioned to support collective migration. Slavery did have long-term effects on individual and institutional outcomes, but this worked through channels that have previously been unexplored and which we will examine in future research.

References

[1] Acemoglu, Daron and James A. Robinson. 2008. Persistence of Power, Elites, and Institutions, *American Economic Review* 98(1):267-293.

[2] Alston, Lee J. and Joseph P. Ferrie. 1999. *Southern Paternalism and the American Welfare State: Economics, Politics, and Institutions in the South, 1865-1965*, Cambridge, MA: Cambridge University Press.

[3] Ballester, Coralio, Antoni Calvò-Armengol, and Yves Zenou. 2006. Who's Who in Networks. Wanted: The Key Player, *Econometrica* 74(5): 1403-1417.

[4] Bernheim, Douglas B., Bezalel Peleg, and Michael D. Whinston. 1987. Coalition-Proof Nash Equilibria. I. Concepts, *Journal of Economic Theory* 42: 1-12.

[5] Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. 2012. The Great Migration and African-American Mortality: Evidence from the Deep South, typescript.

[6] Blassingame, John W. 1972. *The Slave Community: Plantation Life in the Antebellum South*. New York and Oxford: Oxford University Press.

[7] Bloch, Francis. 2005. Group and network formation in Industrial Organization: A survey, in *Group formation in economics: Networks, clubs and coalitions* eds. G. Demange and M. Wooders, Cambridge: Cambridge University Press.

[8] Boles, John B. 1988. Introduction, in *Masters and Slaves in the House of the Lord: Race and Religion in the American South, 1740-1870*, ed. John B. Boles, Lexington: The University Press of Kentucky.

[9] Boustan, Leah Platt. 2010. Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration, *Quarterly Journal of Economics* 125(1): 417-443.

[10] Carrington, William J., Enrica Detragiache, and Tara Vishwanath. 1996. Migration with Endogenous Moving Costs, *American Economic Review* 86(4): 909-930.

[11] Coleman, James S. 1988. Social capital in the creation of human capital, American Journal of Sociology 94(S1): S95-S120.

[12] Collins, William J. 1997. When the Tide Turned: Immigration and the Delay of the Great Black Migration, *Journal of Economic History* 57(3): 607-632.

[13] Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 1999. The Rise and Decline of the American Ghetto, *Journal of Political Economy* 107(3): 455-506.

[14] Du Bois, W.E.B. 1908. *The Negro American Family.* New York: Negro Universities Press.

[15] Dvorak, Katherine L. 1988. After Apocalypse, Moses, in *Masters and Slaves in the House of the Lord: Race and Religion in the American South, 1740-1870*, ed. John B. Boles, Lexington: The University Press of Kentucky.

[16] Earle, Carville. 1992. *Geographical Inquiry and American Historical Problems*, Palo Alto: Stanford University Press.

[17] Elliott, Matthew and Benjamin Golub. 2013. A Network Approach to Public Goods, typescript.

[18] Engerman, Stanley L. and Kenneth L. Sokoloff. 1997. Factor Endowments, Institutions, and Differential Growth Paths Among New World Economies, in *How Latin America Fell Behind*, ed. Stephen Haber, Stanford, CA: Stanford University Press.

[19] Fogel, Robert W. 1989. *Without Consent or Contract: The Rise and Fall of American Slavery.* New York : W.W. Norton and Company.

[20] Fogel, Robert W. and Stanley L. Engerman. 1974. *Time on the Cross: The Economics of American Negro Slavery.* Boston: Little, Brown.

[21] Foner, Eric. 1988. *Reconstruction, America's Unfinished Revolution, 1863-1877.* New York: Harper and Row.

[22] Foner, Eric. 1993. *Freedom's lawmakers: a directory of Black officeholders during Reconstruction.* New York: Oxford University Press, 1993.

[23] Frazier, Franklin E. 1939. *The Negro Family in the United States.* Chicago: University of Chicago Press.

[24] Frazier, Franklin E. 1964. *The Negro Church in America.* Liverpool: Liverpool University Press.

[25] Genicot, Garance and Debraj Ray. 2003. Group Formation in Risk-Sharing Arrangements, *Review of Economic Studies* 70(1): 87-113.

[26] Genovese, Eugene D. 1974. *Roll, Jordan, Roll: The World the Slaves Made.* New York: Pantheon Books.

[27] Goldin, Claudia D. and Frank D. Lewis. 1975. The Economic Cost of the American Civil War: Estimates and Implications, *Journal of Economic History* 35(2): 299-326

[28] Gottlieb, Peter. 1987. *Making Their Own Way: Southern Blacks' Migration to Pittsburgh, 1916-30*. Urbana and Chicago: University of Illinois Press.

[29] Granovetter, Mark S. Economic action and social structure: The problem of embededness, American Journal of Sociology 481-510.

[30] Grossman, James R. 1989. *Land of Hope: Chicago, Black Southerners, and the Great Migration*. Chicago: University of Chicago Press.

[31] Gutman, Herbert G. 1976. *The Black Family in Slavery and Freedom, 1750-1925.* New York: Vintage Books.

[32] Hansen, Bruce E. 1999. Threshold Effects in Non-Dynamic Panels: Estimation, Testing and Inference, *Journal of Econometrics* 93: 345-368.

[33] Henri, Florette. 1975. *Black Migration: Movement North 1900-1920.* Garden City, New York : Anchor Press/Doubleday.

[34] House, Albert W. 1954. Labor Management Problems on Georgia Rice Plantations, 1840-1860, *Agricultural History* 28(4): 149-155

[35] Irwin, James R. 1994. Explaining the Decline in Southern per Capita Output after Emancipation, *Explorations in Economic History* 31: 336-356.

[36] Jackson, Matthew O., Tomas Rodriguez-Barraquer, and Xu Tan, Social Capital and Social Quilts: Network Patterns of Favor Exchange, *American Economic Review* 102(5): 1857-1897.

[37] Karlan, Dean, Markus Mobius, Tanya Rosenblat, and Adam Szeidl. 2009. Trust and Social Capital, *Quarterly Journal of Economics* 124(3): 1307-1361.

[38] King, Ronald F. 2001(a). A Most Corrupt Election: Louisiana in 1876, *Studies in American Political Development* 15:123-137.

[39] King, Ronald F. 2001(b). Counting the Votes: South Carolina's Stolen Election of 1876, *Journal of Interdisciplinary History* 32(2):169-191.

[40] Kolchin, Peter. 1993. *American Slavery, 1619-1877.* New York: Hill and Wang.

[41] Kousser, J. Morgan. 1974. *The Shaping of Southern Politics: Suffrage Restrictions and the Establishment of the One-Party South, 1880-1910*, New Haven and London: Yale University Press.

[42] Lee, Everett S., Ann Ratner Miller, Carol P. Brainerd, and Richard Easterlin. 1957. *Population Redistribution and Economic Growth, United States 1870-1950, Vol. 1: Methodological Considerations and Reference Tables*, Philadelphia: American Philosophical Society.

[43] Lieberson, Stanley. 1978. Selective Migration from the South: A Historical View, in *The Demography of Racial and Ethnic Groups*, eds. Frank D. Bean and W. Parker Frisbie, Academic Press.

[44] Mandle, Jay R. 1978. *The Roots of Black Poverty: The Southern Plantation Economy After the Civil War*. Durham, North Carolina: Duke University Press.

[45] Margo, Robert. 1990. *Race and Schooling in the South, 1880-1950.* Chicago: University of Chicago Press.

[46] Marks, Carole. 1983. Lines of Communication, Recruitment Mechanisms, and the Great Migration of 1916-1918, *Social Problems* 31(1): 73-83.

[47] Marks, Carol. 1989. *Farewell – We're Good and Gone: The Great Black Migration.* Bloomington and Indianapolis: Indiana University Press.

[48] Marks, Carol. 1991. The Social and Economic Life of Southern Blacks During the Migration, in *Black Exodus: The Great Migration from the American South*, ed. Alferdteen Harrison, Jackson and London: University Press of Mississippi.

[49] Moen, Jon R. 1992. Changes in the Productivity of Southern Agriculture between 1860 and 1880, in *Without Consent or Contract: The Rise and Fall of American Slavery; Markets and Production Technical Papers, Volume 1*, eds. Robert William Fogel and Stanley L. Engerman, New York : W.W. Norton and Company.

[50] Morrison, Minion K.C. 1987. *Black Political Mobilization: Leadership, Power, and Mass Behavior*. Albany: State University of New York Press.

[51] Munshi, Kaivan. 2003. Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market, *Quarterly Journal of Economics* 118(2): 549-597.

[52] Naidu, Suresh. 2010. Recruitment Restrictions and Labor Markets: Evidence from the Postbellum U.S. South, *Journal of Labor Economics* 28(2): 413-445.

[53] Newey, Whitney K., James L. Powell and Francis Vella. 1999. Nonparametric Estimation of Triangular Simultaneous Equations Models, *Econometrica* 67(3): 565-603.

[54] Niles Weekly Register. 1835. "Culture of Sugar Cane" from the New Orleans Bulletin, Volume 49: 129 (October 23).

[55] Olstead, Alan and Paul W. Rhode. 2010. Slave Productivity in Cotton Production by Gender, Age, Season, and Scale, University of California, Davis, typescript.

[56] Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.

[57] Robinson, Peter. 1988. "Root-N-Consistent Semiparametric Regression," *Econometrica* 56: 931-954.

[58] Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. Integrated Public Use Microdata Series (IPUMS): Version 5.0. Minneapolis: University of Minnesota.

[59] Stampp, Kenneth M. 1956. *The Peculiar Institution: Slavery in the Ante-Bellum South*. New York: Alfred A. Knopf.

[60] Stampp, Kenneth M. 1966. *The Era of Reconstruction, 1865-1877*. New York: Alfred A. Knopf.

[61] Steckel, Richard H. 2000. The African American Population of the United States, 1790-1920, in *A Population History of North America*, eds. Michael Haines and Richard H. Steckel, New York: Cambridge University Press.

[62] Tadman, Michael. 1989. *Speculators and Slaves: Masters, Traders, and Slaves in the Old South*. Madison: University of Wisconsin Press.

[63] Tolnay, Stewart E. and E.M. Beck. 1990. Lethal Violence and the Great Migration, 1900-1930, *Social Science History* 14(3): 347-370.

[64] Tolnay, Stuart E. 1998. Educational Selection in the Migration of Southern Blacks, 1880-1990, *Social Forces* 77(3): 487-514.

[65] Valelly, Richard M. 2004. *The Two Reconstructions*. Chicago and London: University of Chicago Press.

[66] Wilson, William Julius. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago: University of Chicago Press.

[67] Woodson, Carter G. 1921. *The History of the Negro Church*. Washington, D.C.: The Associated Publishers.

[68] Wright, Gavin. 1978. *The Political Economy of the Cotton South: Households, Markets, and Wealth in the Nineteenth Century*, New York: W.W. Norton and Company.

[69] Wright, Gavin. 1986. *Old South, New South: Revolutions in the Southern Economy Since the Civil War*, New York: Basic Books, Inc.

Figure 1: Plantation Share Across Southern Counties in 1890

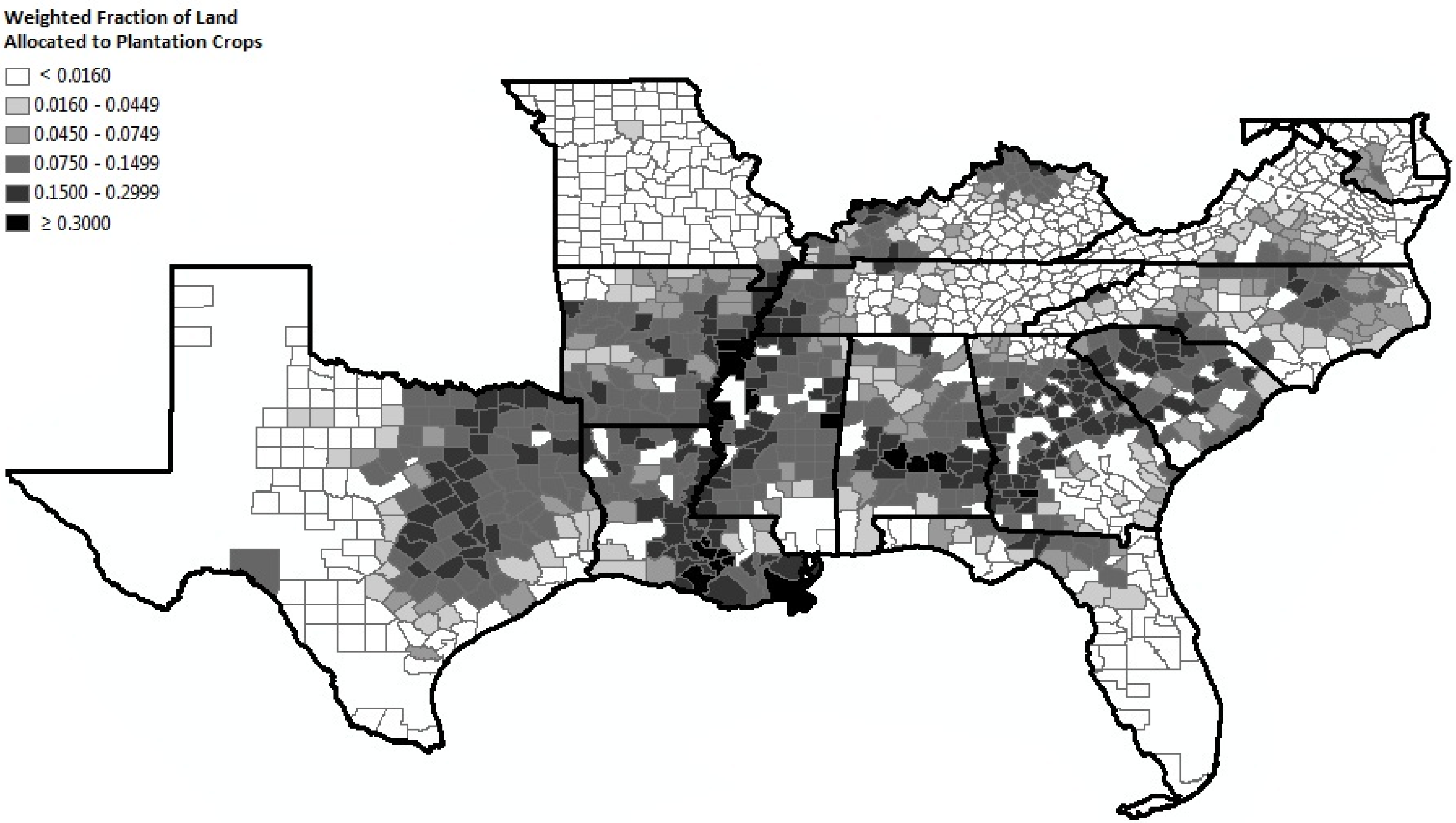
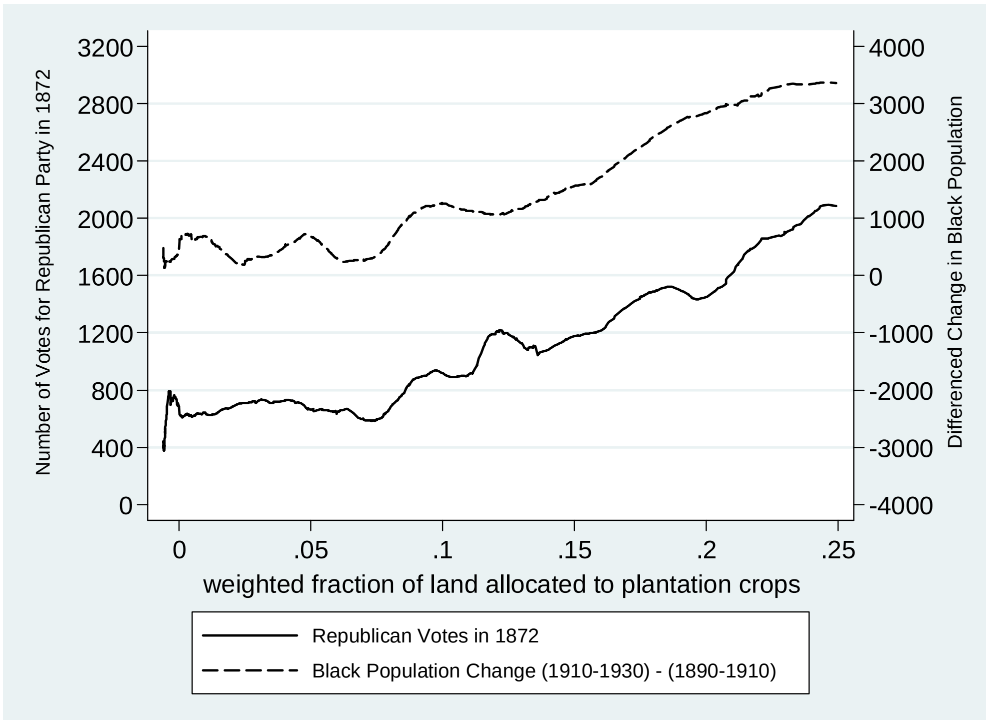


Figure 2: Response to Political and Economic Opportunities

A. Black Population Change and Republican Votes in 1872



B. Migration from Mississippi

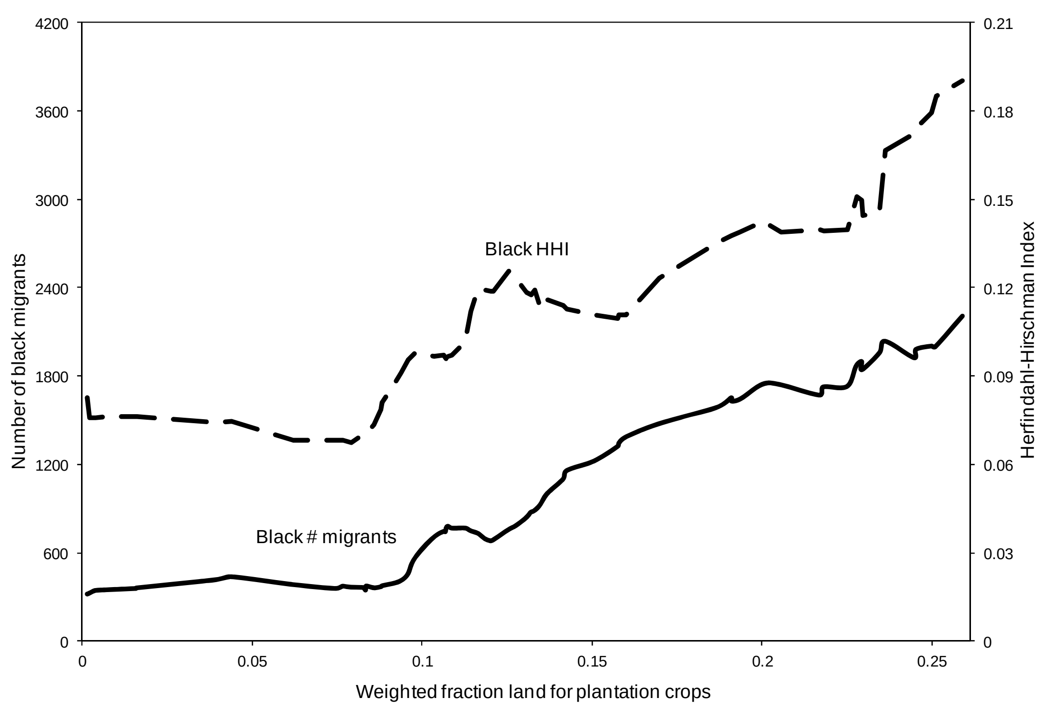
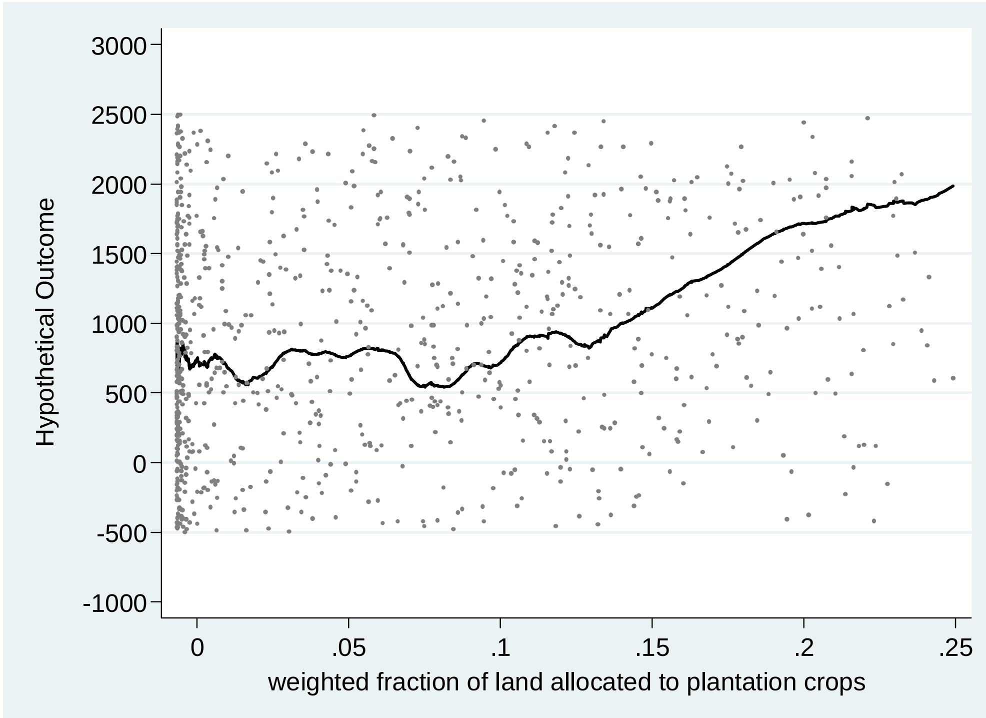
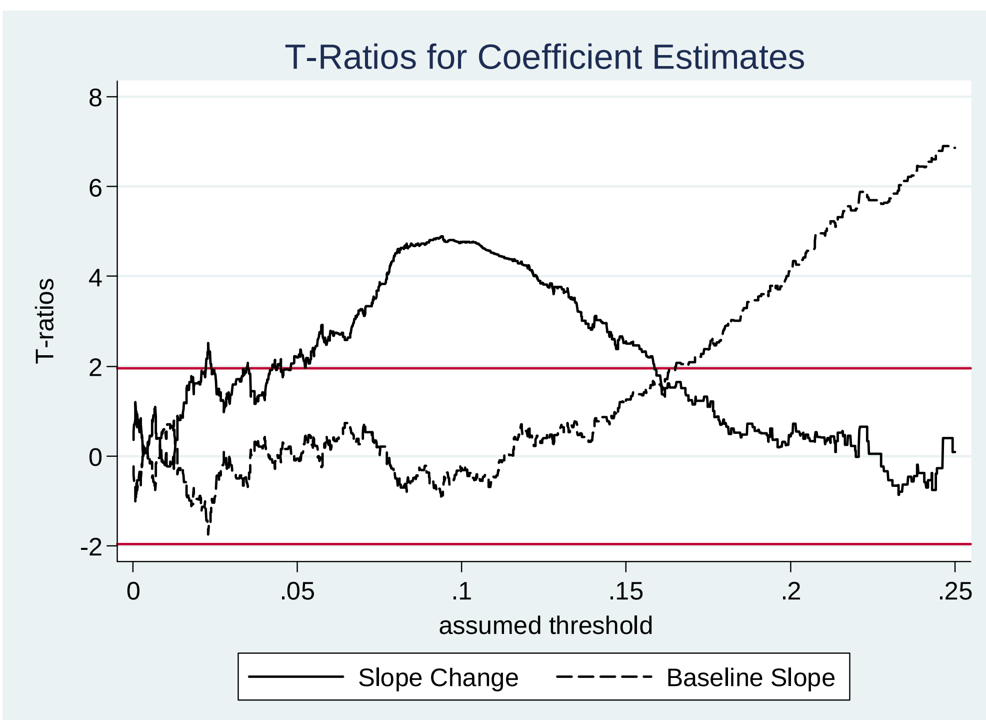


Figure 3: Model Simulation

A. Relationship between hypothetical outcome and plantation share

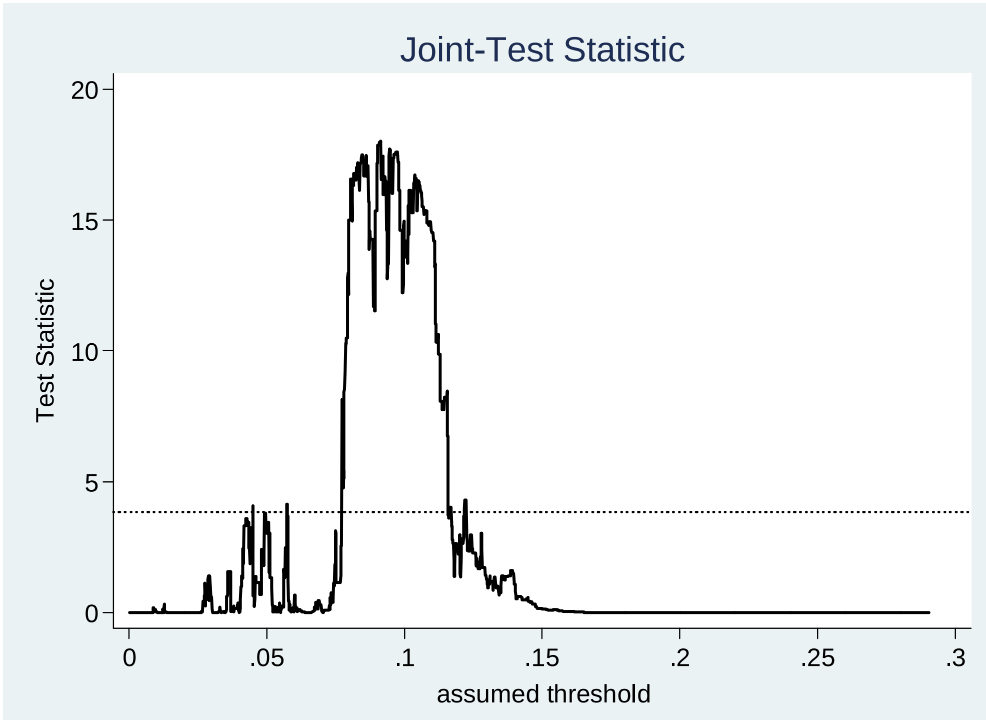


B. T-ratios



\

C. Joint-Test



D. Hansen Test

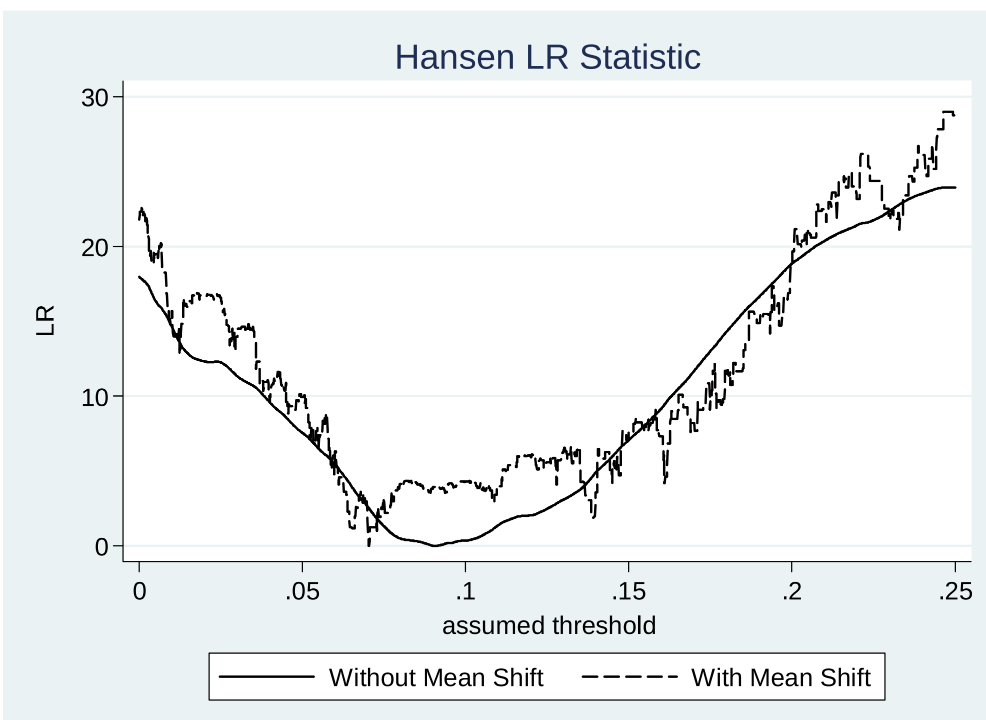
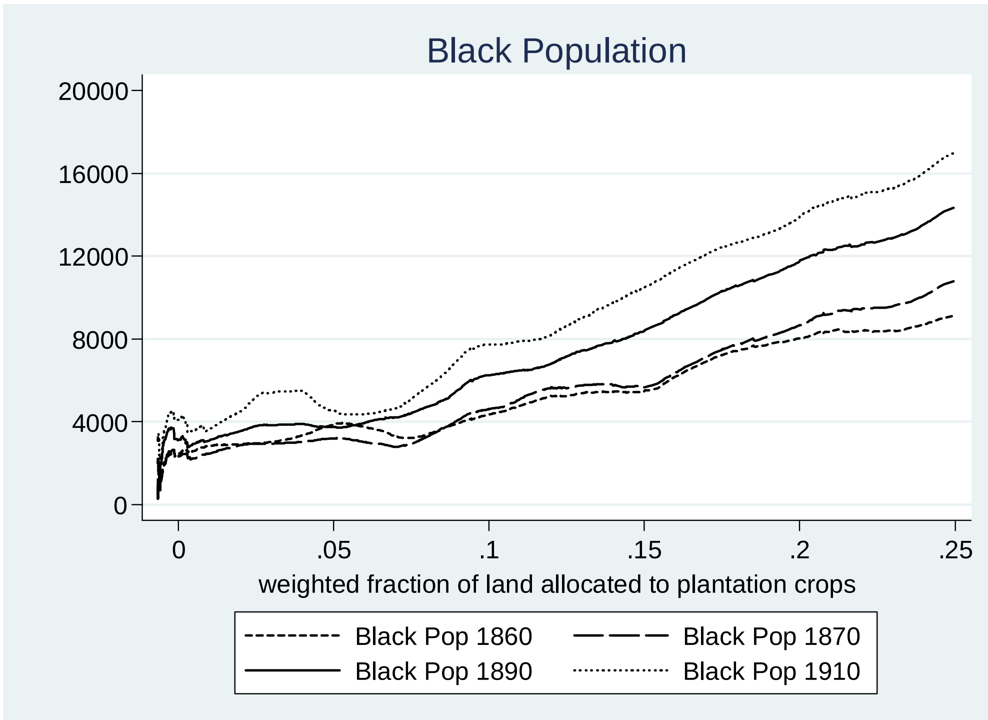
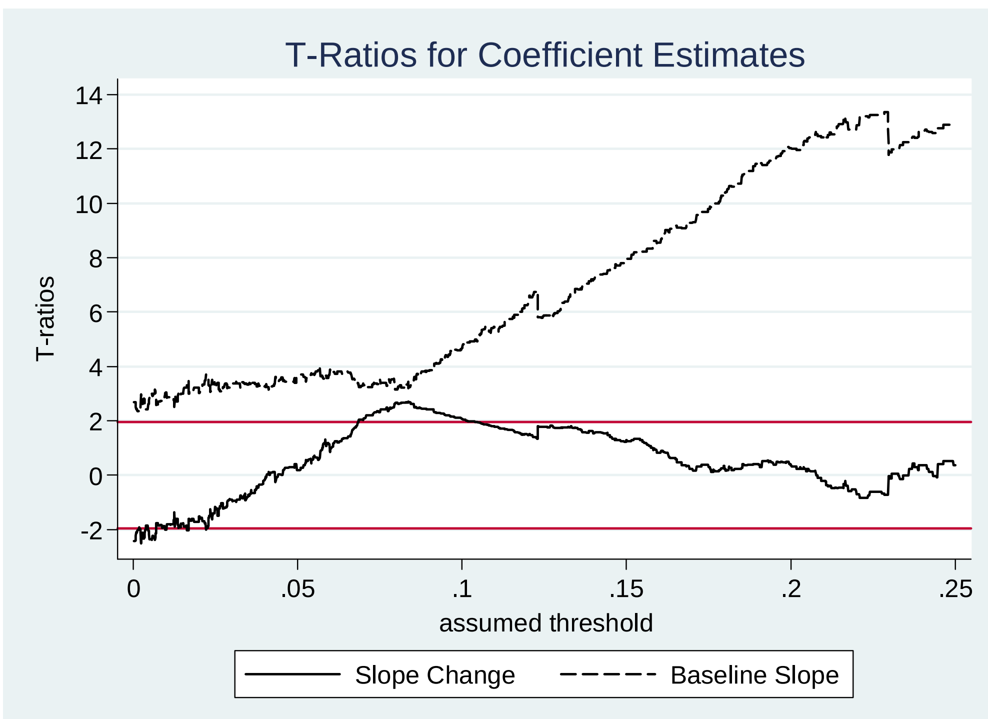


Figure 4: Black Population

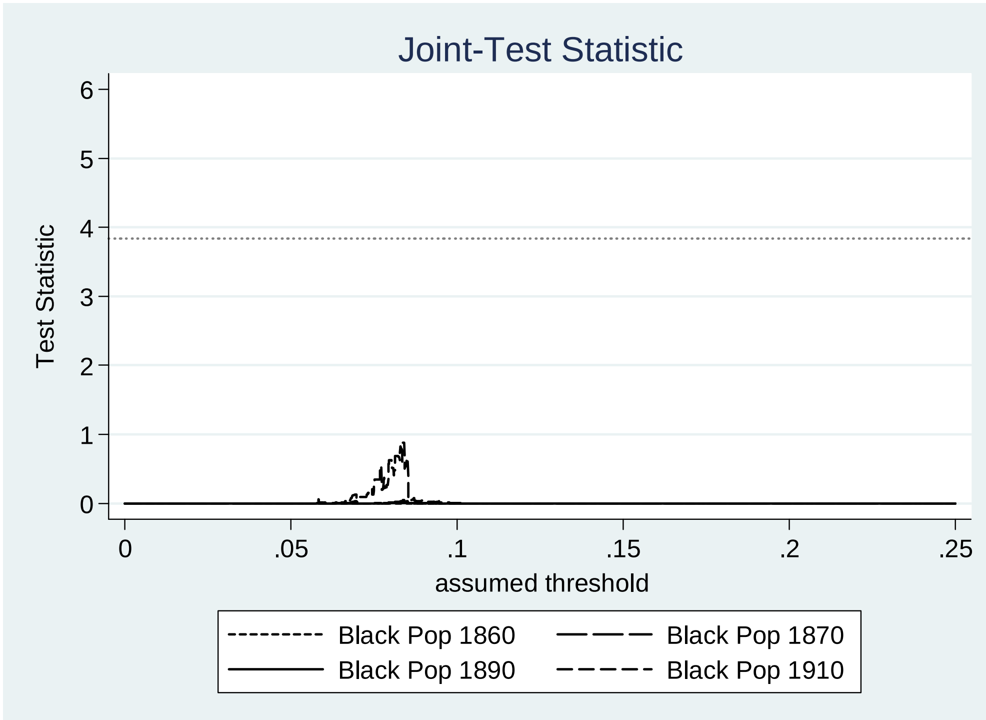
A. Black Population 1860 to 1910



B. T-ratios for Black Population in 1870



C. Joint-Test Statistic for Black Population 1860 to 1910



D. Instrumental Variable Estimates for Black Population in 1870

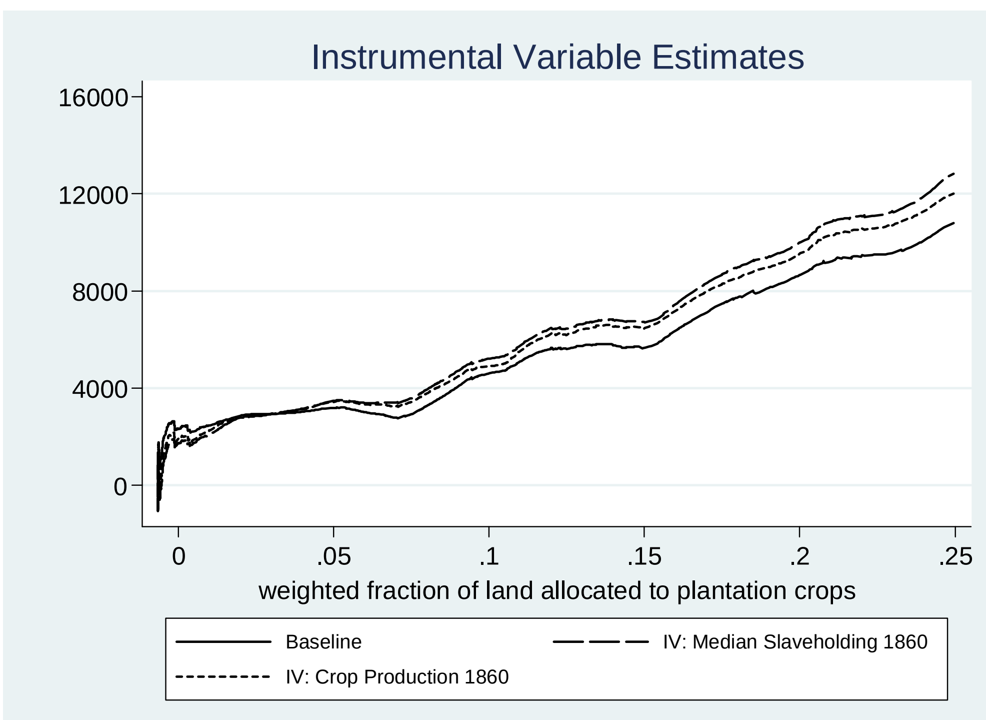
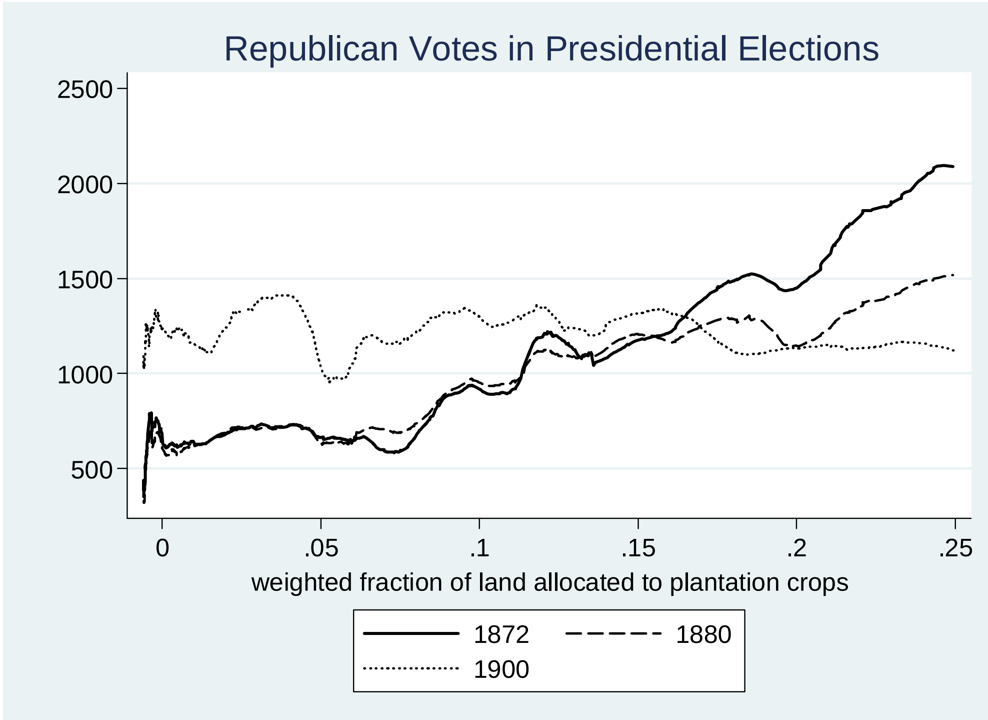
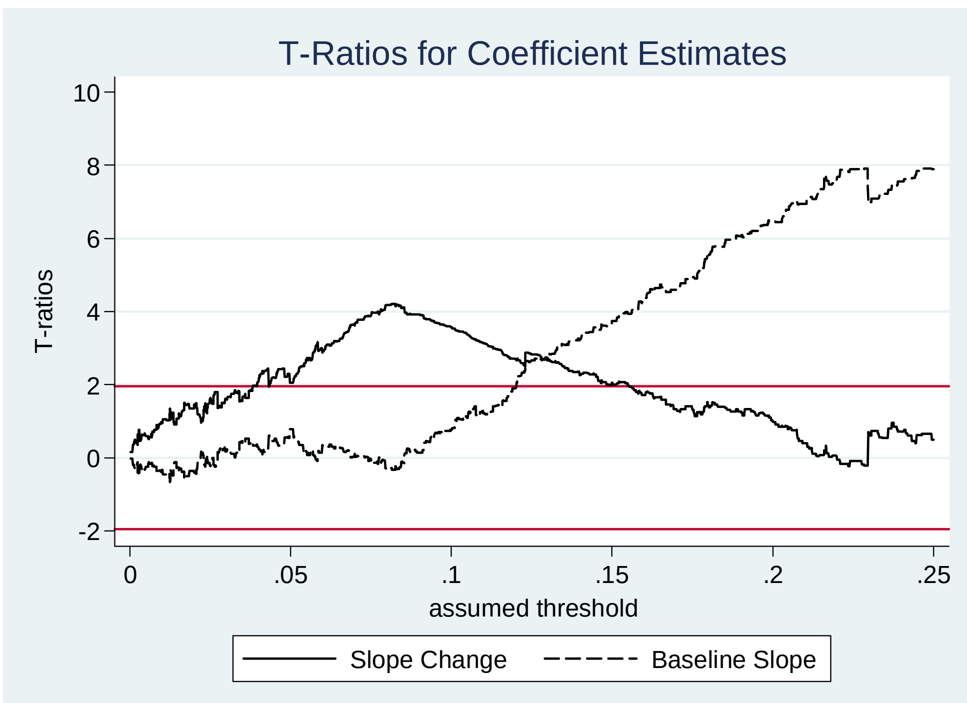


Figure 5: Political Participation

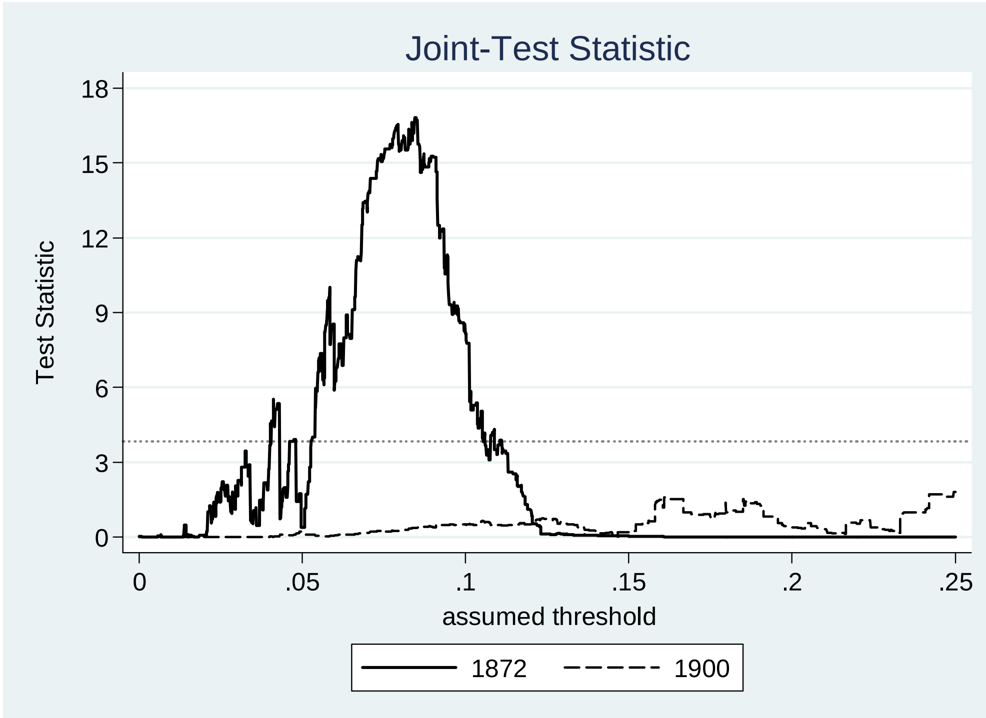
A. Republican Votes in Presidential Elections



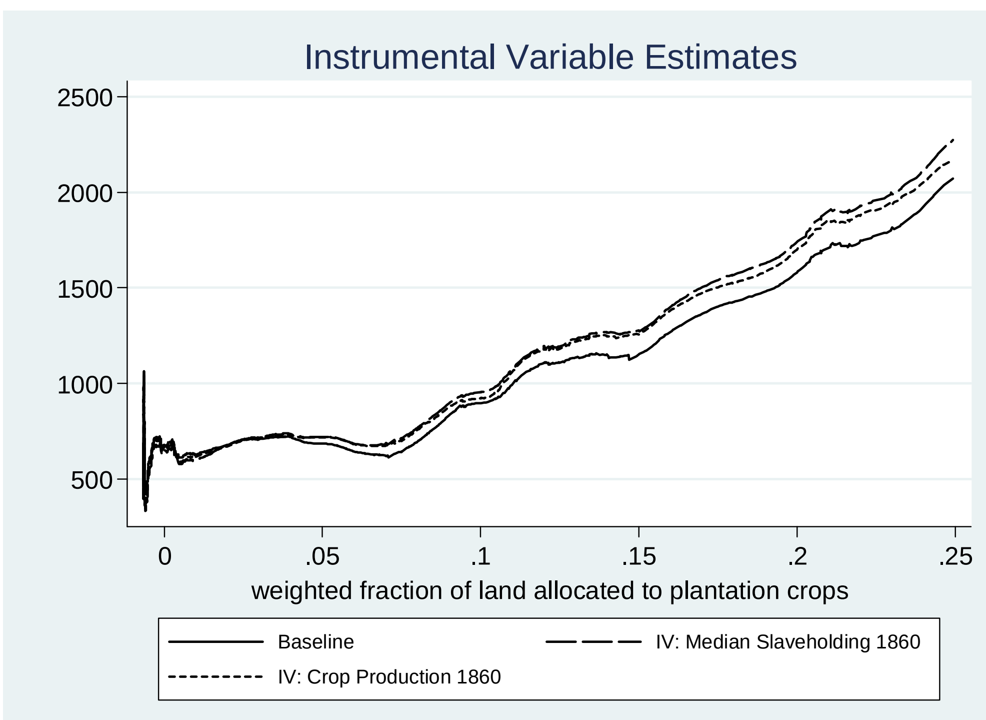
B. T-ratios for Republican Votes in 1872



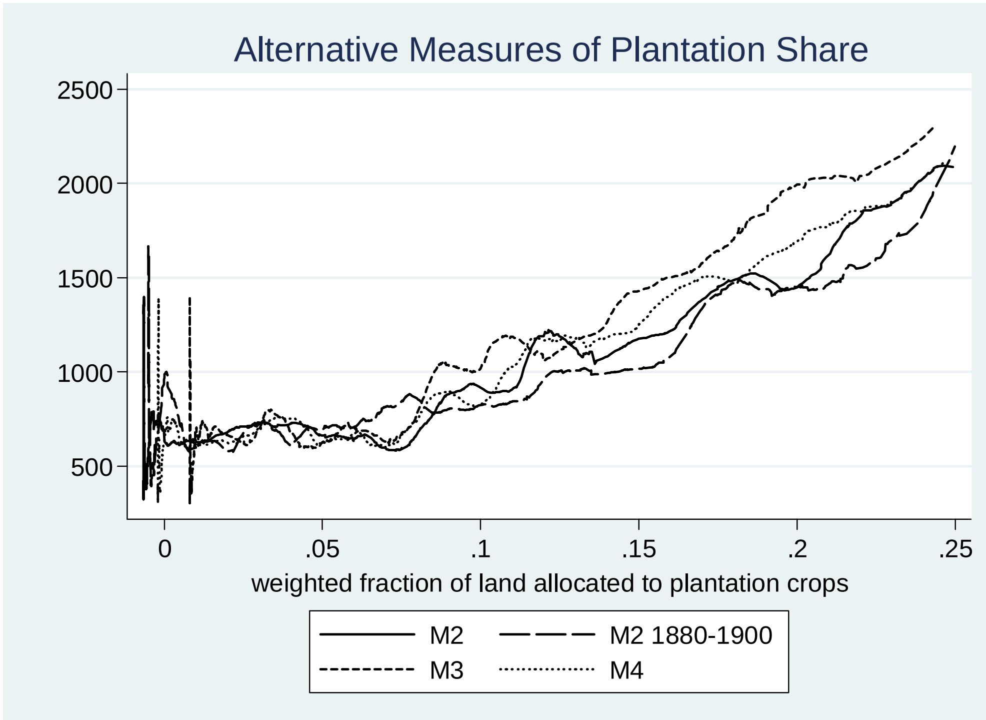
C. Joint-Test for Republican Votes in 1872



D. Instrumental Variable Estimates for Republican Votes in 1872



E. Robustness to Alternative Measures of Plantation Share



F. Congressional and Gubernatorial Elections

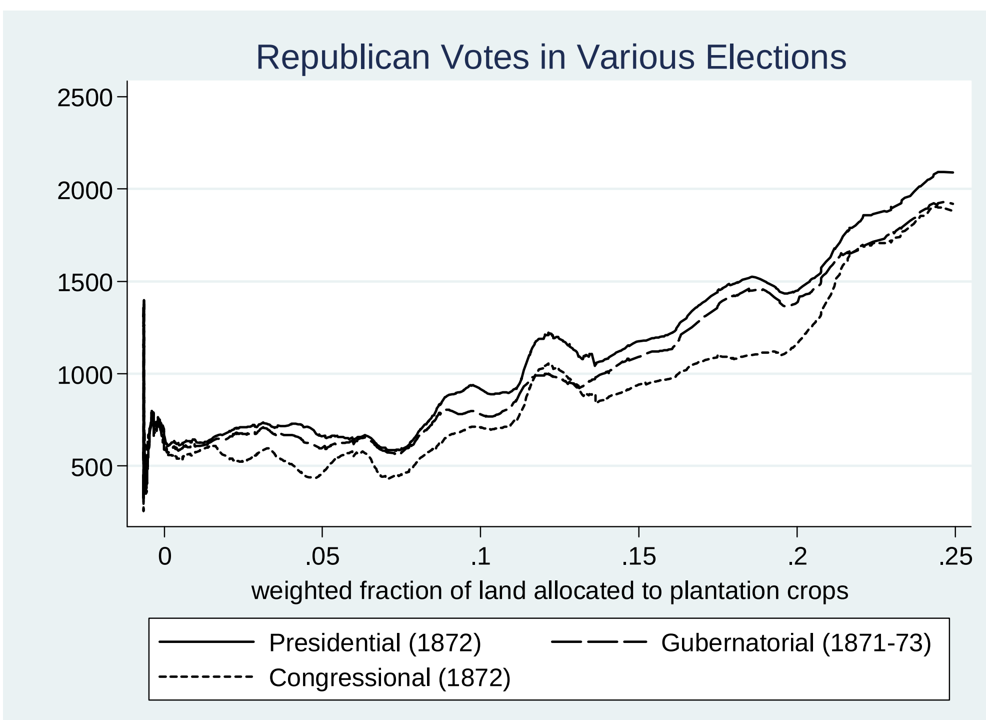
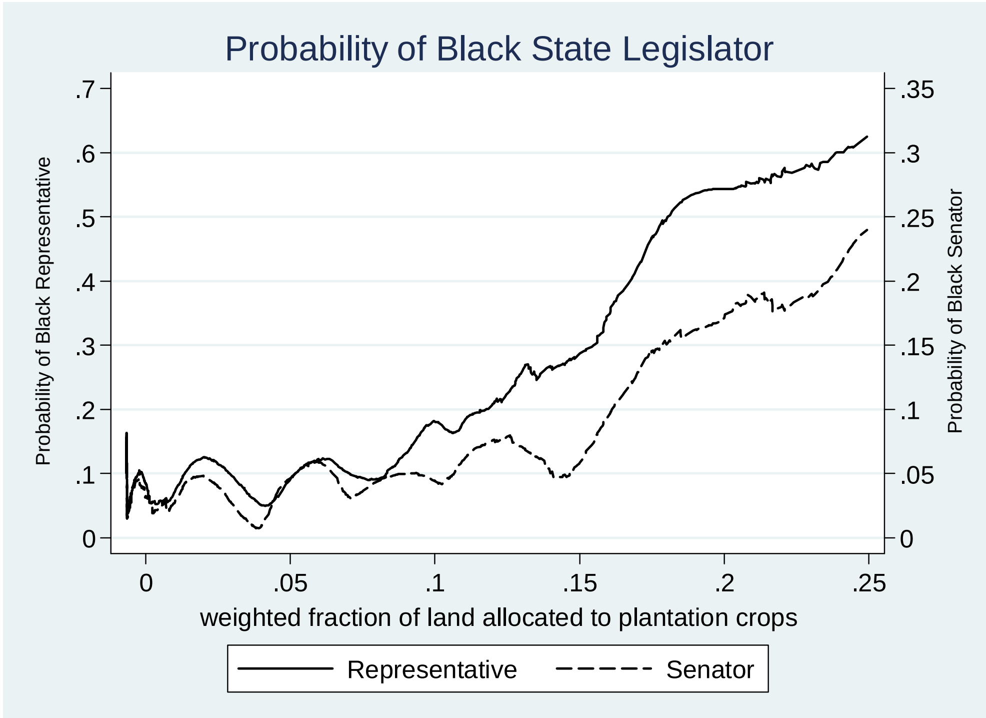
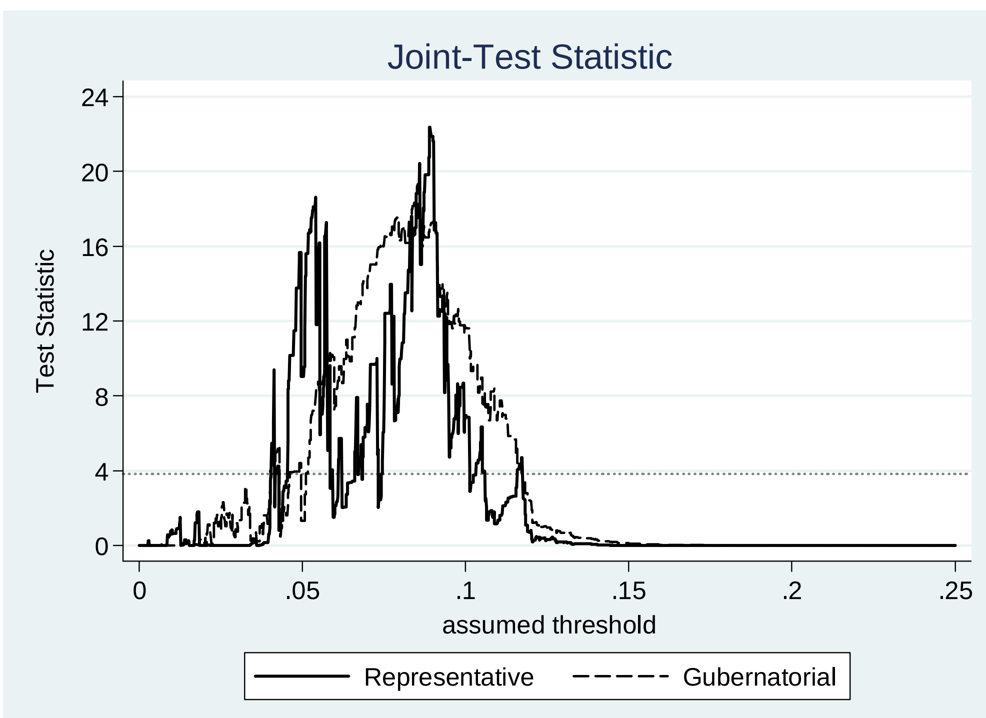


Figure 6: Elected Black Leaders

A. Black State Representative and Senator



B. Joint-Test for Republican Votes in Gubernatorial Elections and Probability of Black State Representative



C. Conditional Estimates for Black State Representative

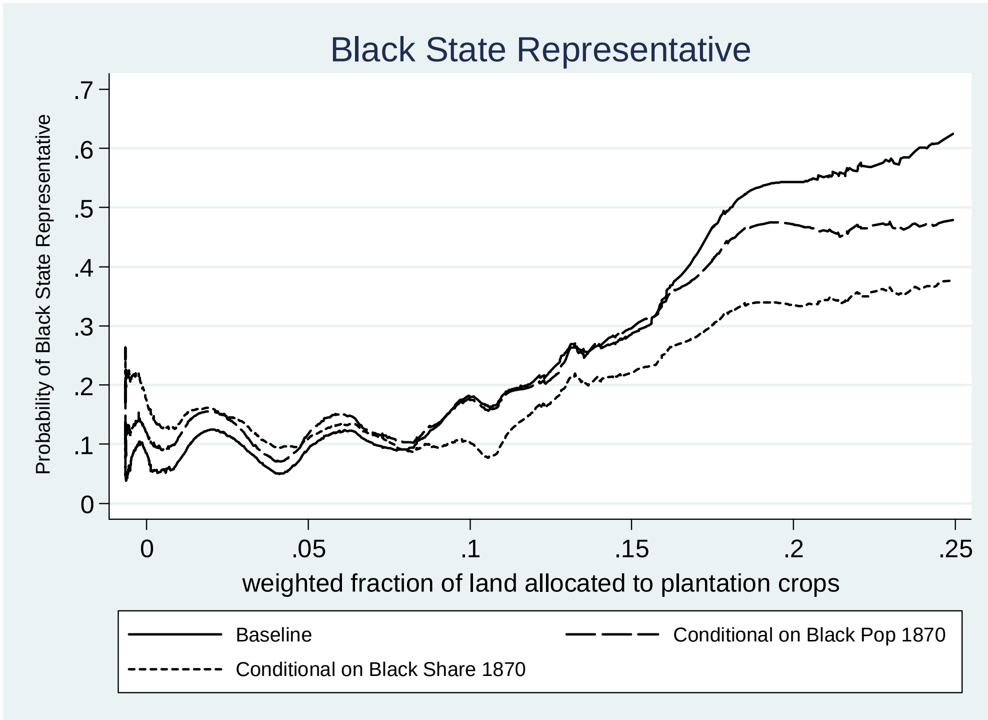
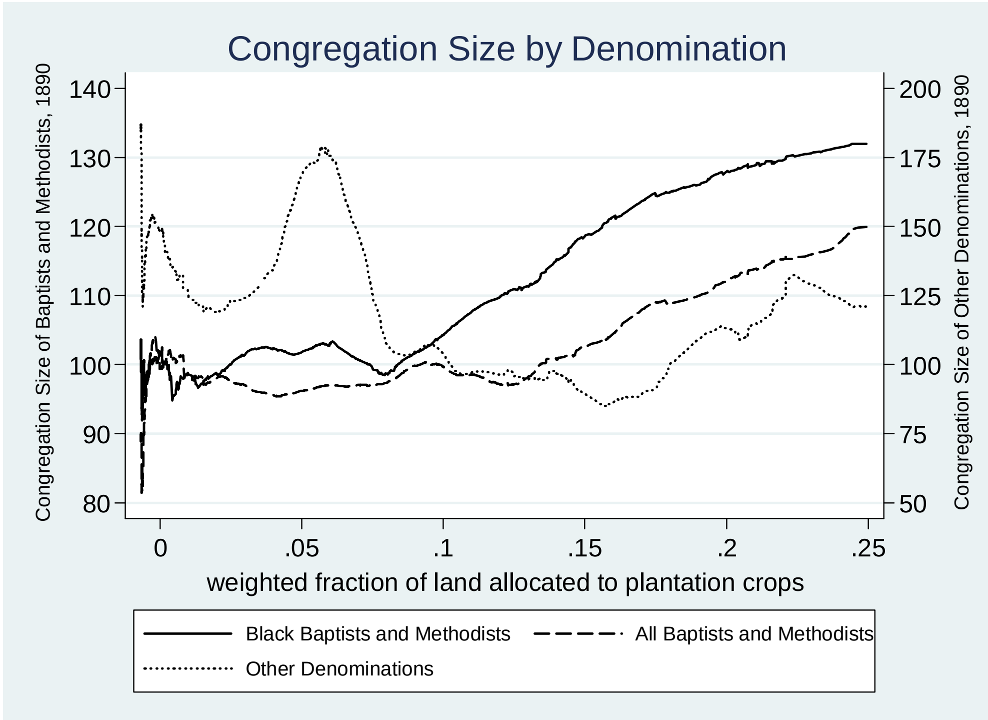
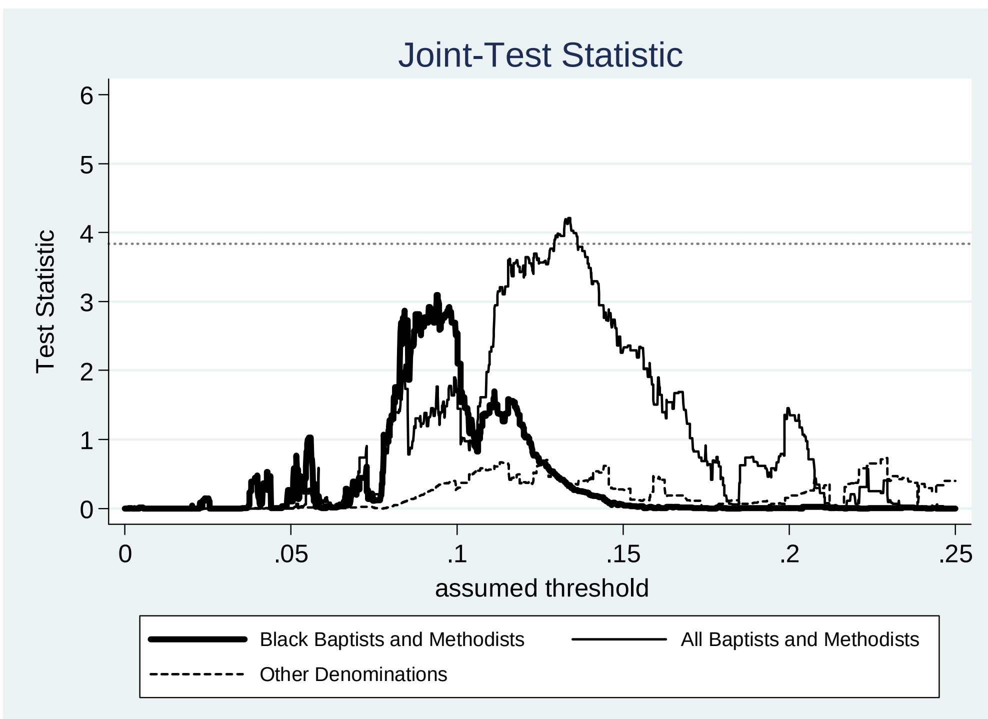


Figure 7: Church Congregation Size in 1890

A. Congregation Size by Denomination



B. Joint-Test for Congregation Size by Denomination



C. Instrumental Variable Estimates for Black Baptist and Methodist Church Size

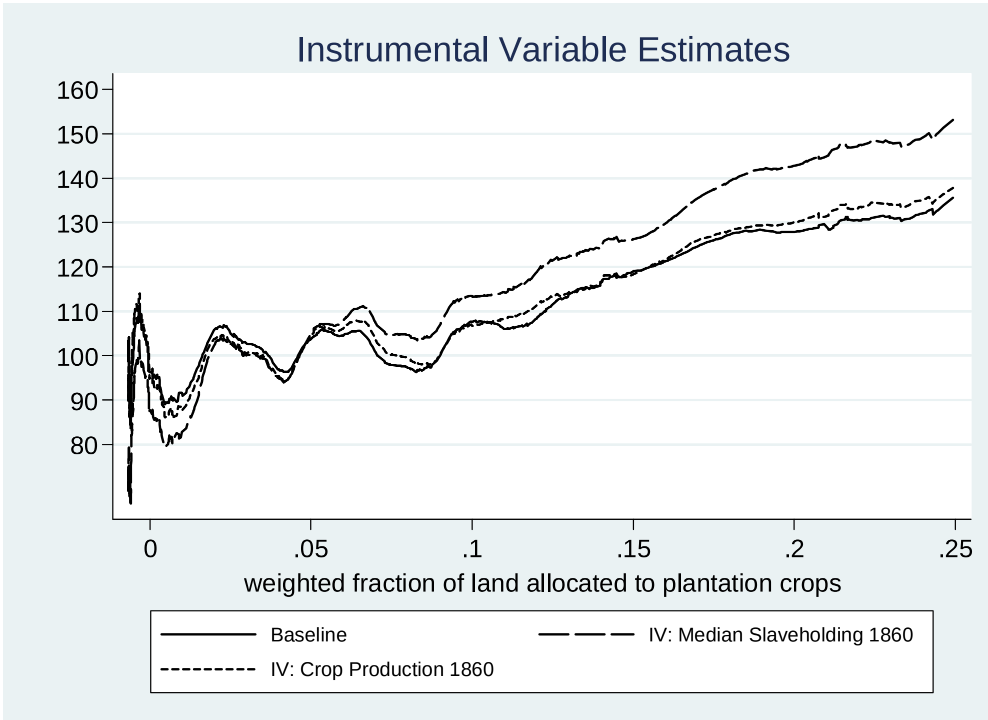
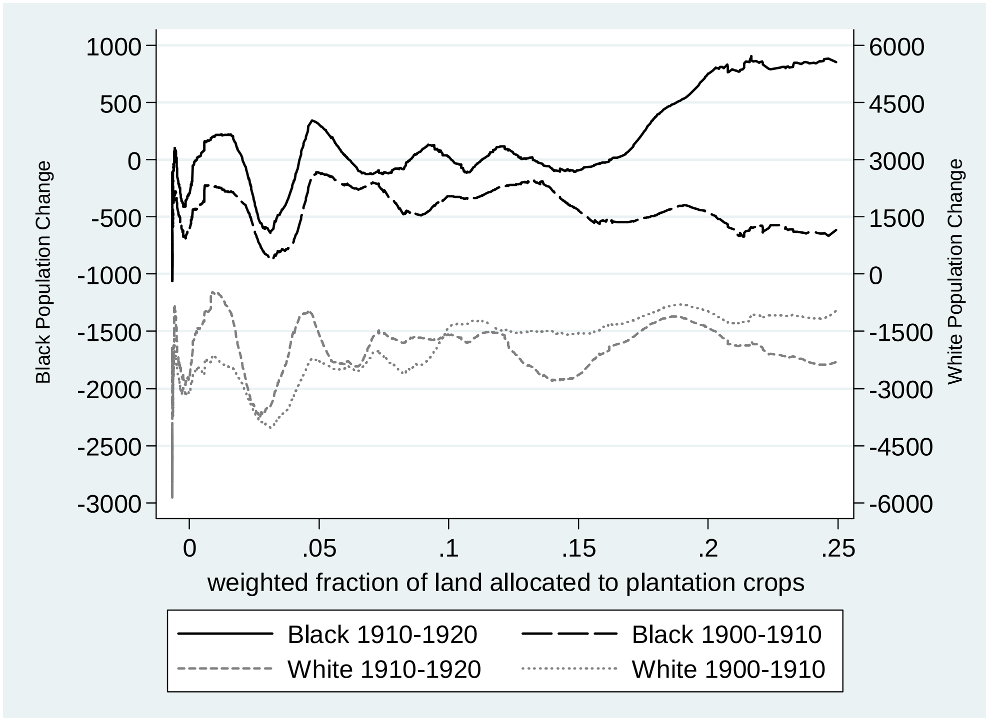
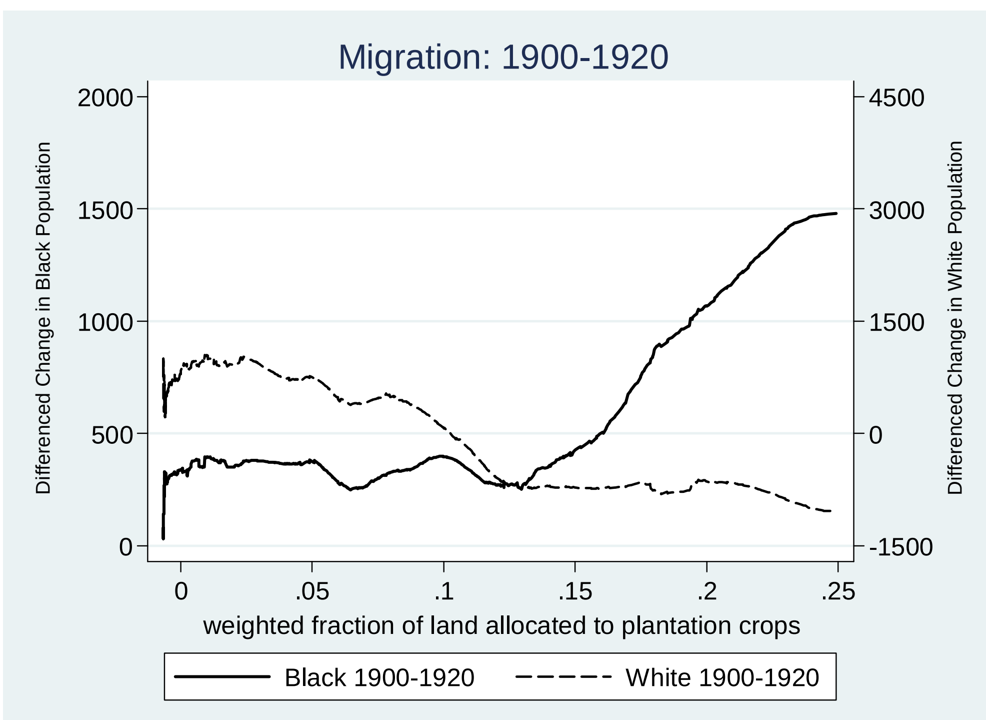


Figure 8: Migration

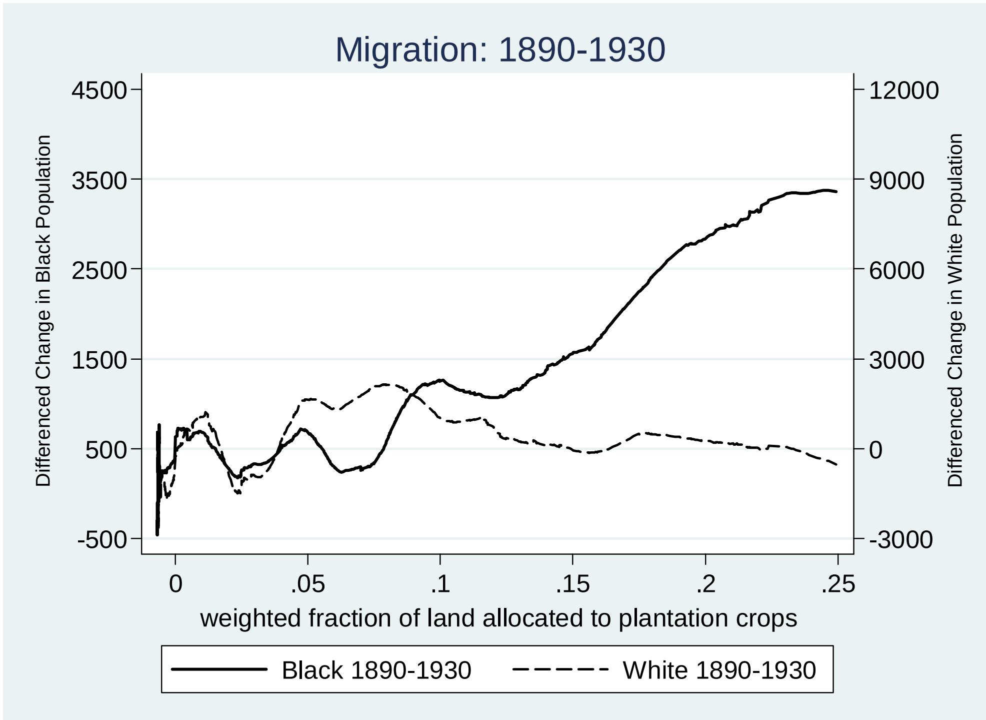
A. Population Change



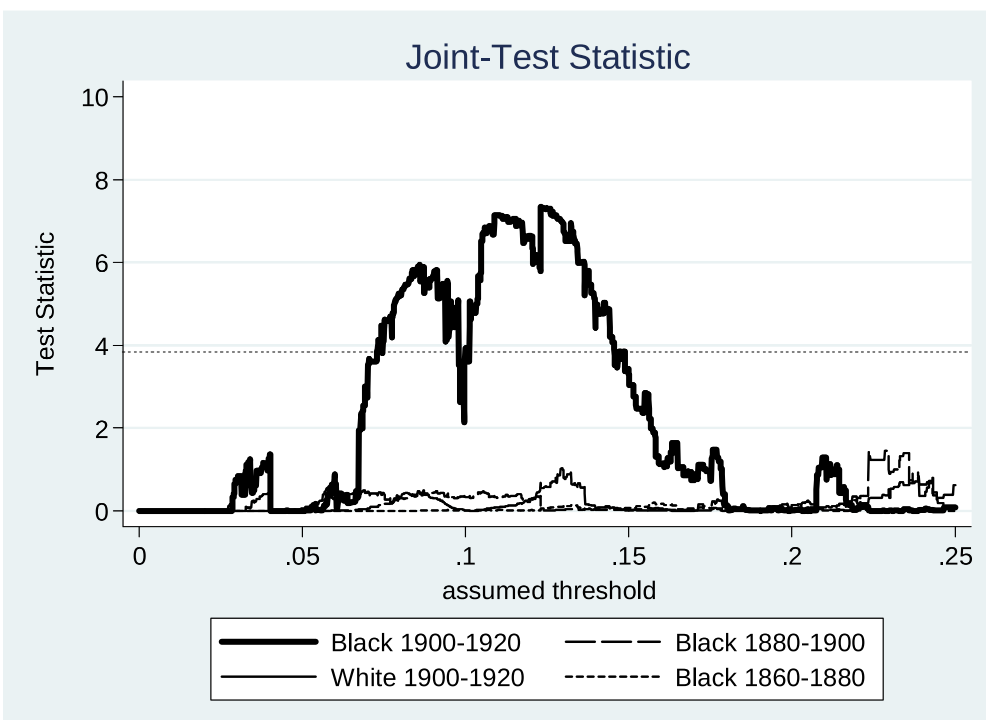
B. Short Double-Difference in Black and White Population



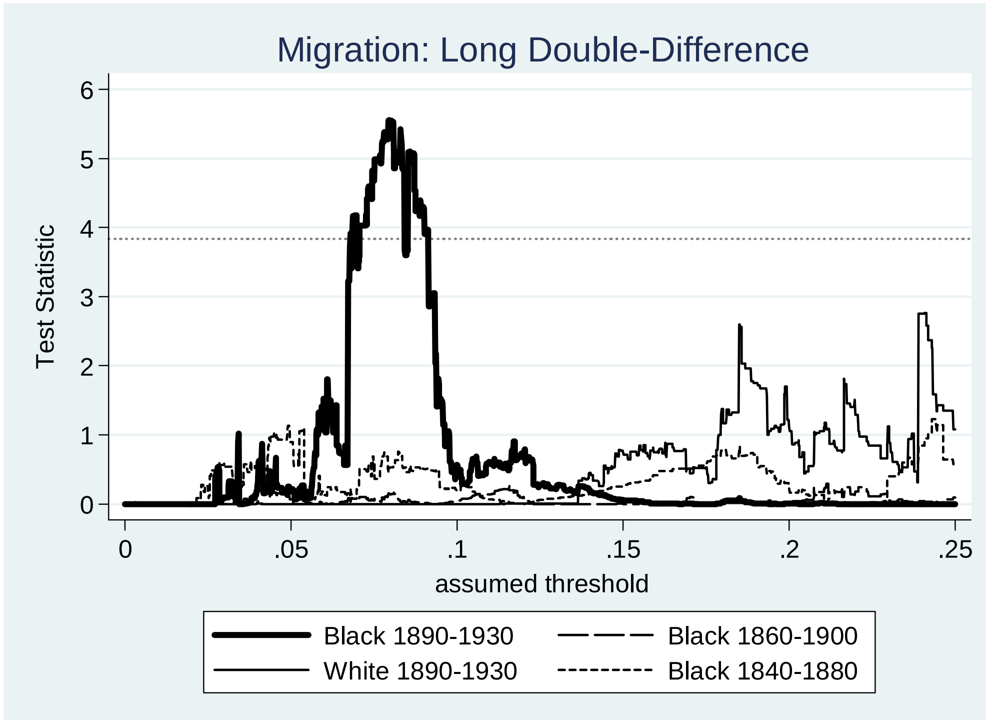
C. Long Double-Difference in Black and White Population



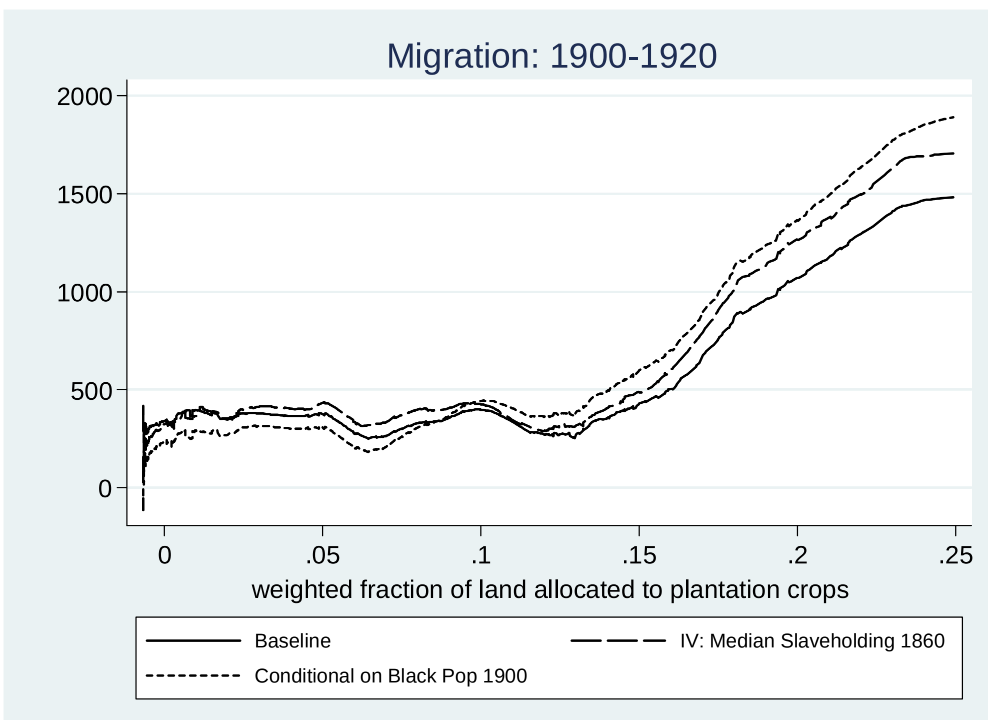
D. Joint-Test for Short Double-Difference in Black and White Population



E. Joint-Test for Long Double-Difference in Black and White Population



F. Instrumental Variable and Conditional Estimates for Short Double-Difference in Black Population



G. Instrumental Variable and Conditional Estimates for Long Double-Difference in Black Population

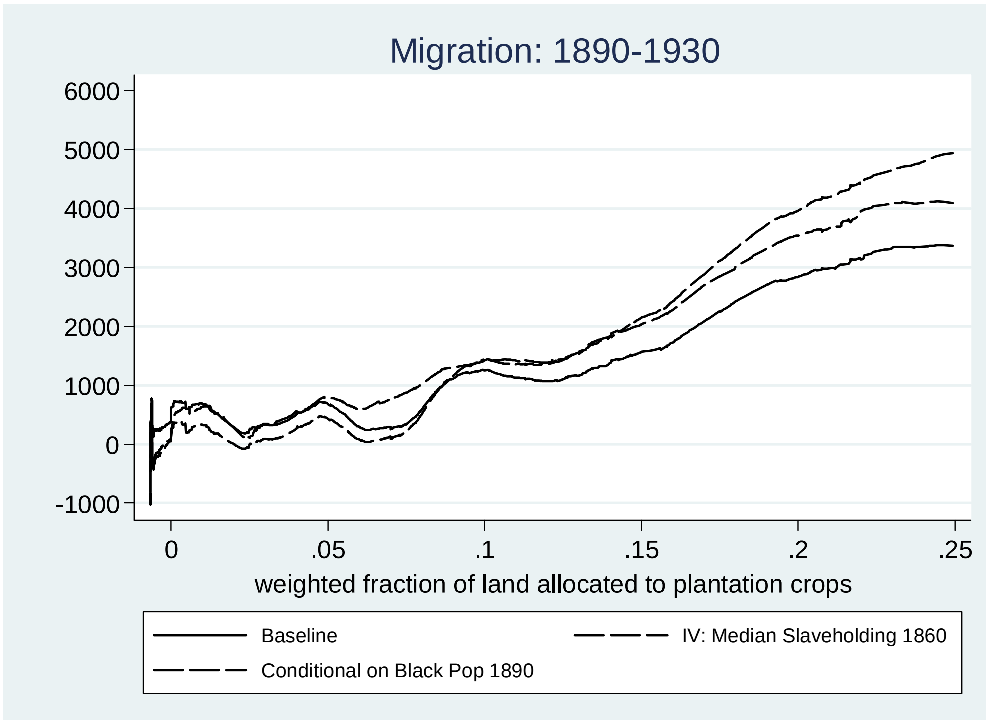
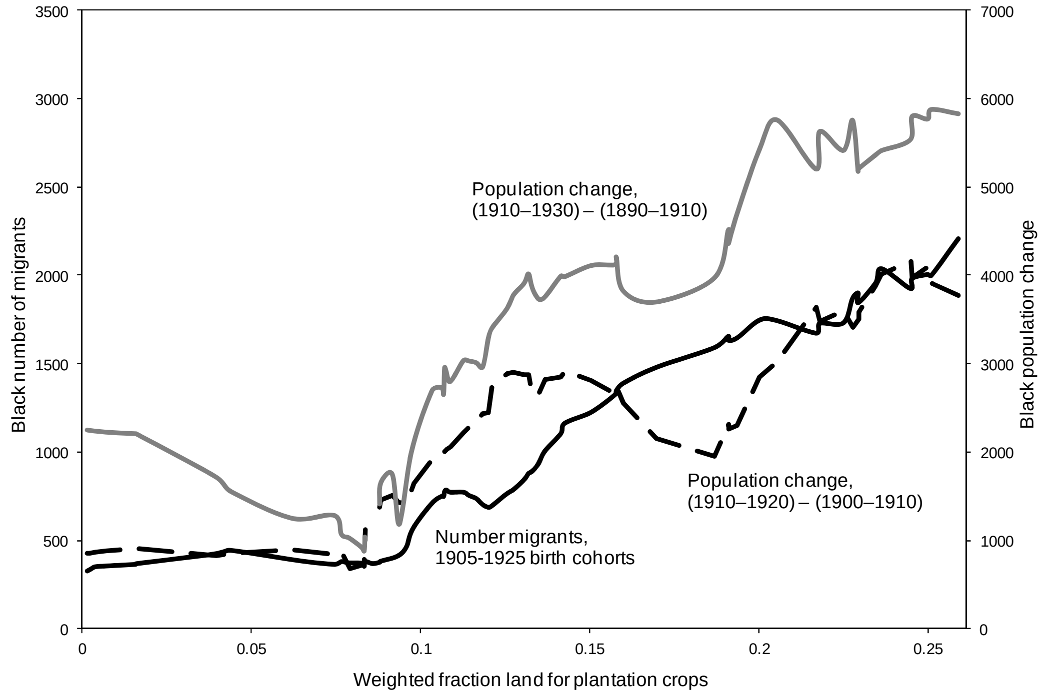


Figure 9: Migration from Mississippi

A. Alternative Measures of Black Migration



B. Level and Distribution of Migrants

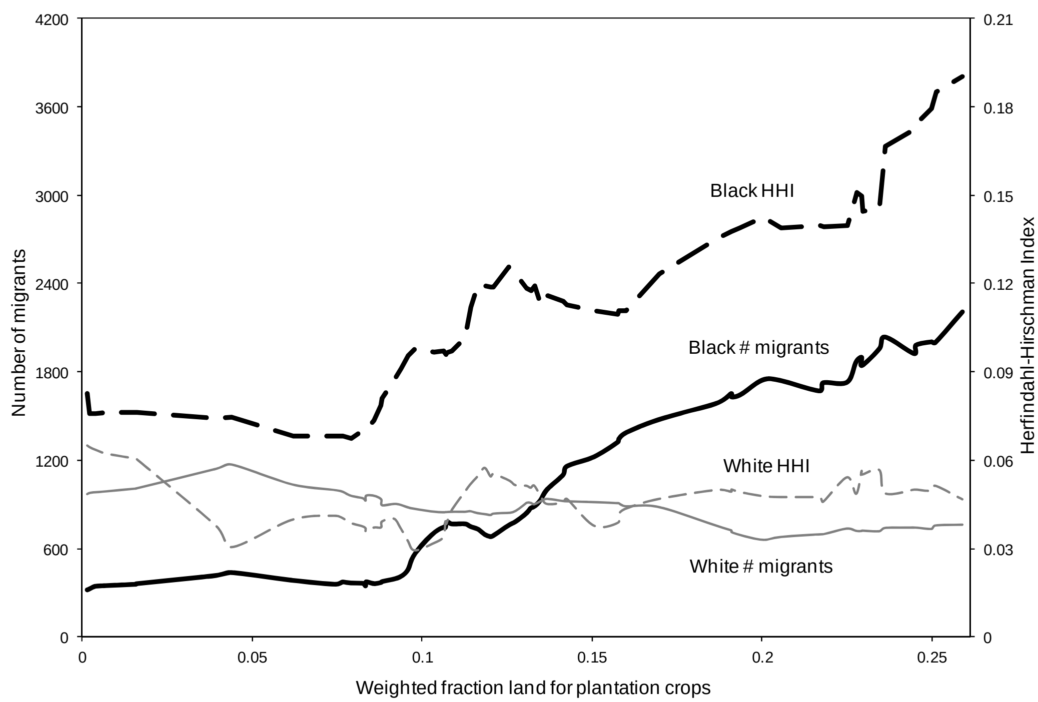


Figure 10: Alternative Explanations

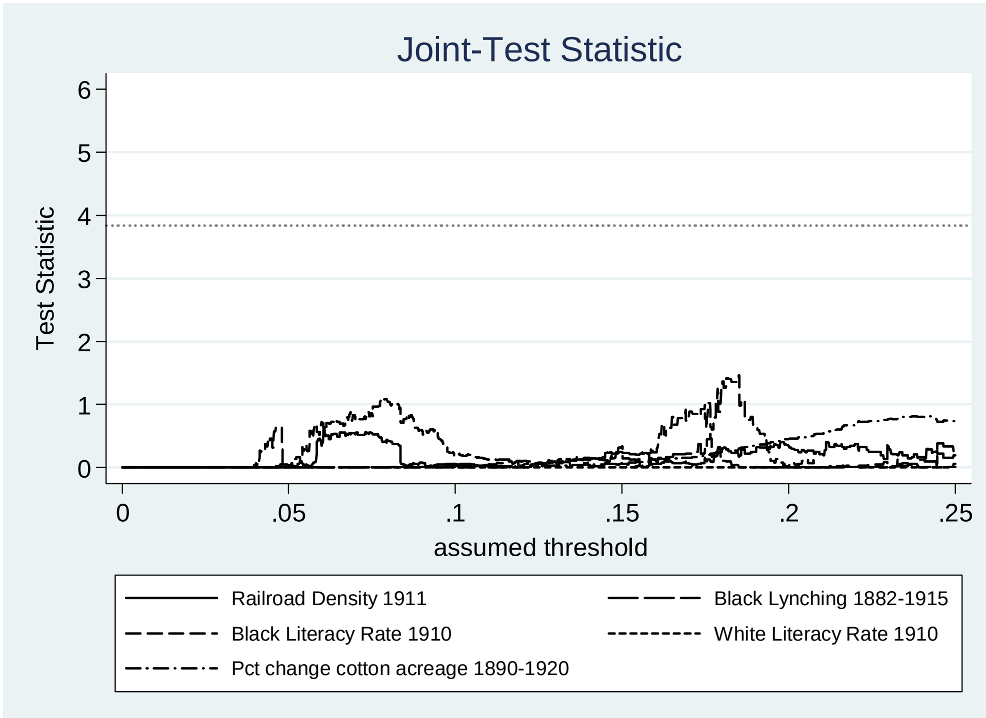


Table 1: Crop Labor Intensities

Density Measure:	M2	M3	M4
Dependent Variable:	---	Black population 1880-1900	Agriculture related occupations 1880
	(1)	(2)	(3)
$\hat{\beta}_{COTTON}$	0.08	0.06***	0.03***
		(0.01)	(0.002)
$\hat{\beta}_{SUGARCANE}$	0.13	0.29***	0.06***
		(0.11)	(0.01)
$\hat{\beta}_{RICE}$	0.15	0.10*	0.14***
		(0.06)	(0.05)
$\hat{\beta}_{TOBACCO}$	0.33	0.16***	0.11***
		(0.02)	(0.01)
$\hat{\alpha}$	---	0.0003	-0.0004
		(0.0004)	(0.0003)
R-squared	---	0.283	0.598
Number of Observations		3361	1133

Notes: Labor intensities in Column 1 are obtained from Olstead and Rhodes (2010), Niles Weekly Register (1835), House (1954), and Earle (1992), respectively.
Estimated labor intensities in Column 2 are obtained by regressing black population on crop acreage, 1880-1900, with county fixed-effects.
Agriculture related occupations in Column 3 include farm laborers, farmers (owners and tenants), and other laborers.
Standard errors in parenthesis are corrected for heteroskedasticity in Columns 2-3 and are also clustered within county in Column 2.

Table 2: Regression results for network outcomes, thresholds determined by joint test statistic [absolute value of t-ratios]

	Republican votes for President in 1872	Republican votes for Governor in 1871-1873	Black State Represent.	Baptist & Methodist church size in 1890	Black Bapt. & Methodist church size in 1890	Diff-in-diffs in Black population, 1900-1920	Diff-in-diffs in Black population, 1890-1930
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Slope change	7688.4***	7659.8***	3.040***	179.48***	193.75*	11136.0***	15579.7**
	[4.10]	[4.25]	[5.53]	[2.61]	[1.78]	[2.74]	[2.36]
Baseline slope	-135.9	-60.9	0.227	41.87	9.61	341.8	-54.5
	[0.10]	[0.05]	[0.61]	[0.94]	[0.10]	[0.23]	[0.01]
Threshold mean shift	118.8	32.5	0.034	-3.18	6.76	-187.8	337.4
	[1.06]	[0.33]	[0.81]	[0.73]	[1.06]	[0.68]	[0.72]
Avg. state fixed effect	667.58	637.04	0.076	94.54	99.06	298.77	382.85
Threshold location	0.0845	0.0843	0.0891	0.1323	0.0940	0.1235	0.0797
Asymptotic p-value of joint test	0.001	0.001	0.001	0.057	0.082	0.008	0.019
R-squared	0.2095	0.1854	0.3384	0.1416	0.2215	0.1299	0.1434
Sample size	1040	1005	1135	1104	939	1126	1125

Notes: Threshold locations based on the maximum joint test statistic. Models include state indicators in the regression, and the estimated standard errors are corrected for heteroskedasticity.

Table 3: Regression results for network outcomes, thresholds determined by Hansen test statistic [absolute value of t-ratios]

A. Without mean shift at threshold

	Republican votes for President in 1872	Republican votes for Governor in 1871-1873	Black State Represent.	Baptist & Methodist church size in 1890	Black Bapt. & Methodist church size in 1890	Diff-in-diffs in Black population, 1900-1920	Diff-in-diffs in Black population, 1890-1930
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Slope change	7856.9***	7821.6***	2.975***	162.06**	194.52*	11312.0***	15539.3**
	[3.94]	[4.16]	[5.24]	[2.28]	[1.73]	[2.73]	[2.27]
Baseline slope	-39.25	-262.91	0.347	32.43	35.0	42.66	1010.2
	[0.03]	[0.24]	[1.06]	[0.83]	[0.41]	[0.04]	[0.25]
Threshold location	0.0741	0.0780	0.0830	0.1328	0.0828	0.1364	0.0700
Bootstrapped p-value of slope change	0.000	0.000	0.000	0.005	0.058	0.002	0.030
R-squared	0.075	0.079	0.166	0.015	0.023	0.021	0.037
Sample size	1036	1001	1129	1104	939	1126	1125

Notes: Threshold locations based on the Hansen test statistic for a structural break without a mean shift included in the regression model. Data has been deviated from state-specific means, and the estimated standard errors are corrected for heteroskedasticity.

B. With mean shift at threshold

	Republican votes for President in 1872	Republican votes for Governor in 1871-1873	Black State Represent.	Baptist & Methodist church size in 1890	Black Bapt. & Methodist church size in 1890	Diff-in-diffs in Black population, 1900-1920	Diff-in-diffs in Black population, 1890-1930
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Slope change	7812.4***	7839.6***	2.979***	161.81**	196.84	11617.3**	14780.6**
	[3.92]	[4.23]	[5.40]	[2.41]	[1.61]	[2.45]	[2.20]
Baseline slope	27.36	-293.51	0.339	32.31	30.18	193.9	2111.7
	[0.02]	[0.20]	[0.83]	[0.73]	[0.26]	[0.16]	[0.39]
Threshold mean shift	-6.95	3.543	0.0011	0.043	0.602	-55.16	-110.6
	[0.07]	[0.04]	[0.03]	[0.01]	[0.09]	[0.17]	[0.24]
Threshold location	0.0741	0.0780	0.0830	0.1328	0.0828	0.1364	0.0700
R-squared	0.075	0.079	0.166	0.015	0.023	0.021	0.037
Sample size	1036	1001	1129	1104	939	1126	1125

Notes: Threshold locations based on the Hansen test statistic for a structural break without a mean shift included in the regression model. Data has been deviated from state-specific means, and the estimated standard errors are corrected for heteroskedasticity.

Appendix

Figure A1: Short Double-Difference in Black Population Prior to Great Migration

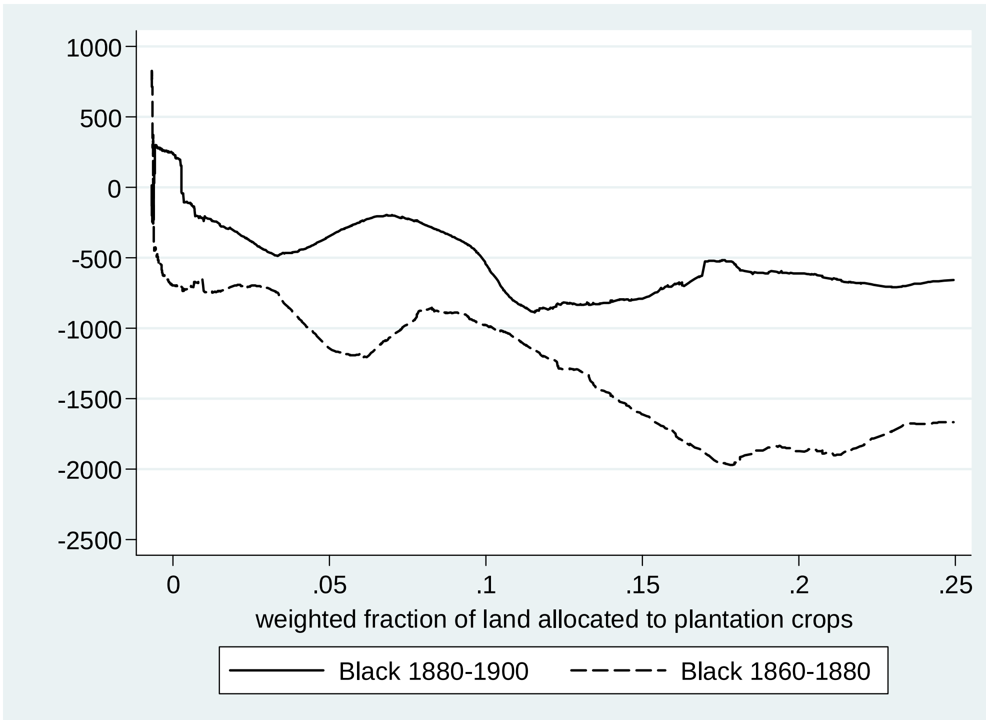
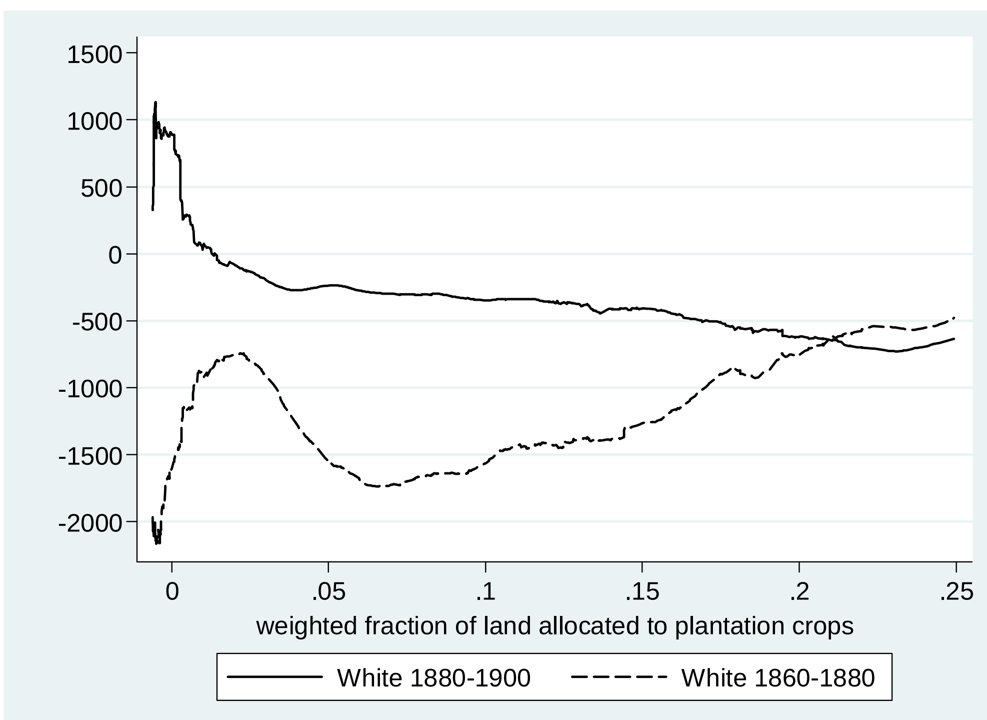


Figure A2: Short Double-Difference in White Population Prior to Great Migration



\

Figure A3: Long Double-Difference in Black Population Prior to Great Migration

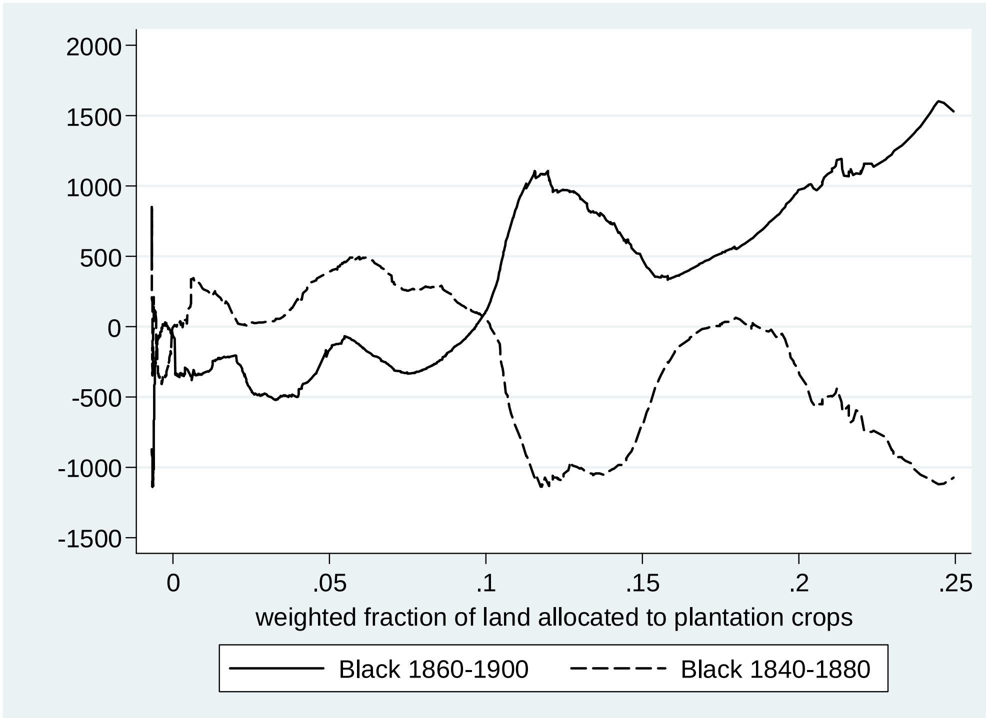


Figure A4: Long Double-Difference in White Population Prior to Great Migration

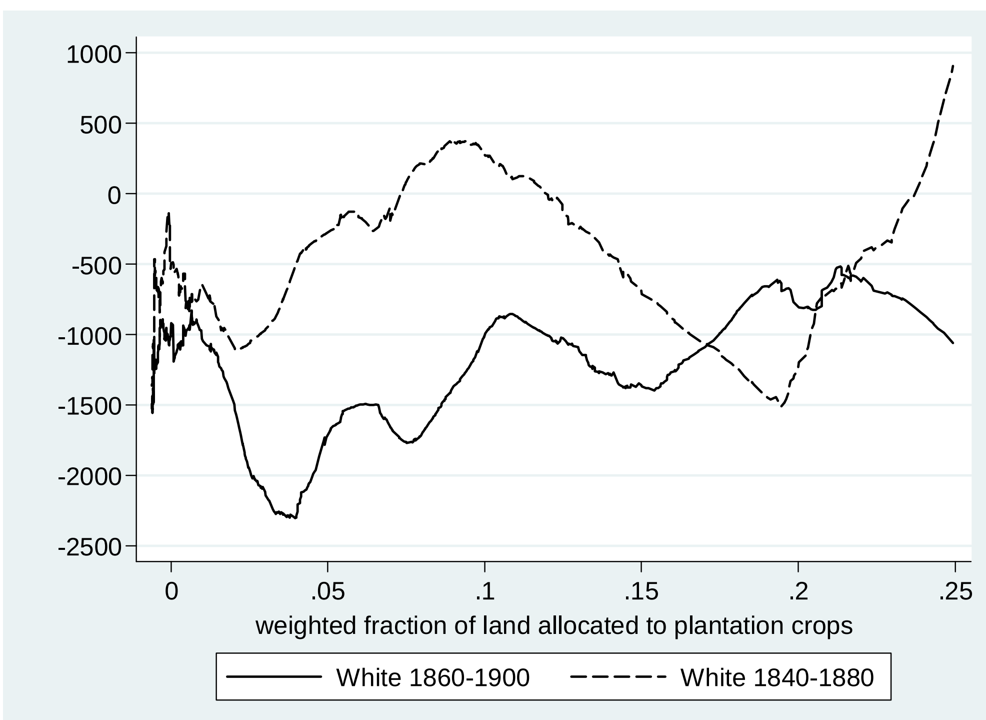


Figure A5: Hansen Test for Republican Votes in Various Elections and Probability of Black State Representative

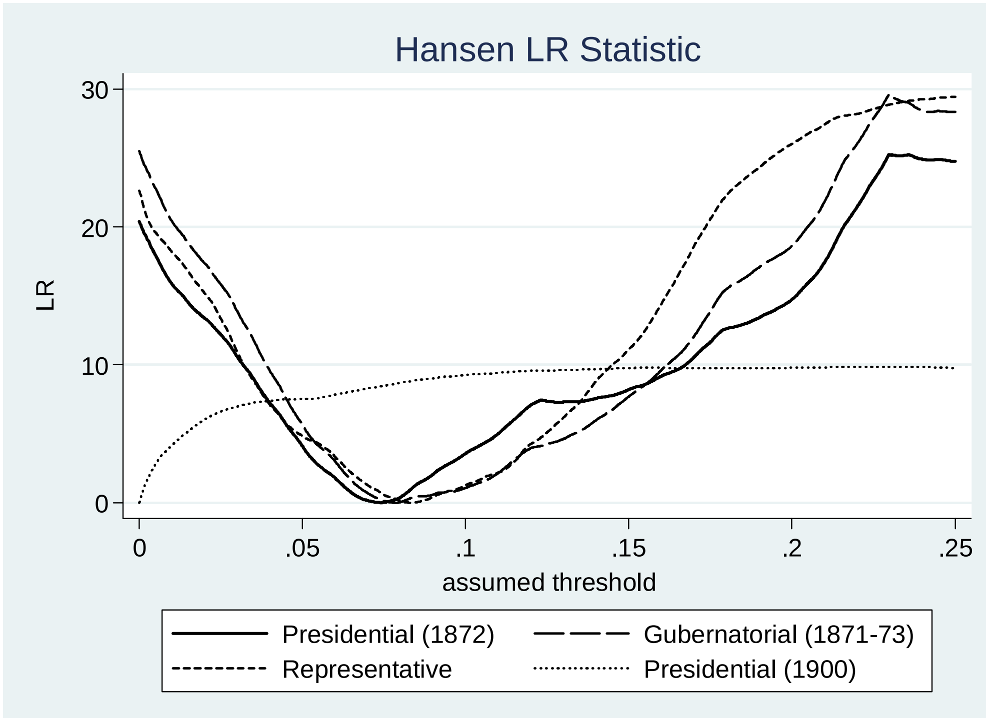


Figure A6: Hansen Test for Congregation Size by Denomination

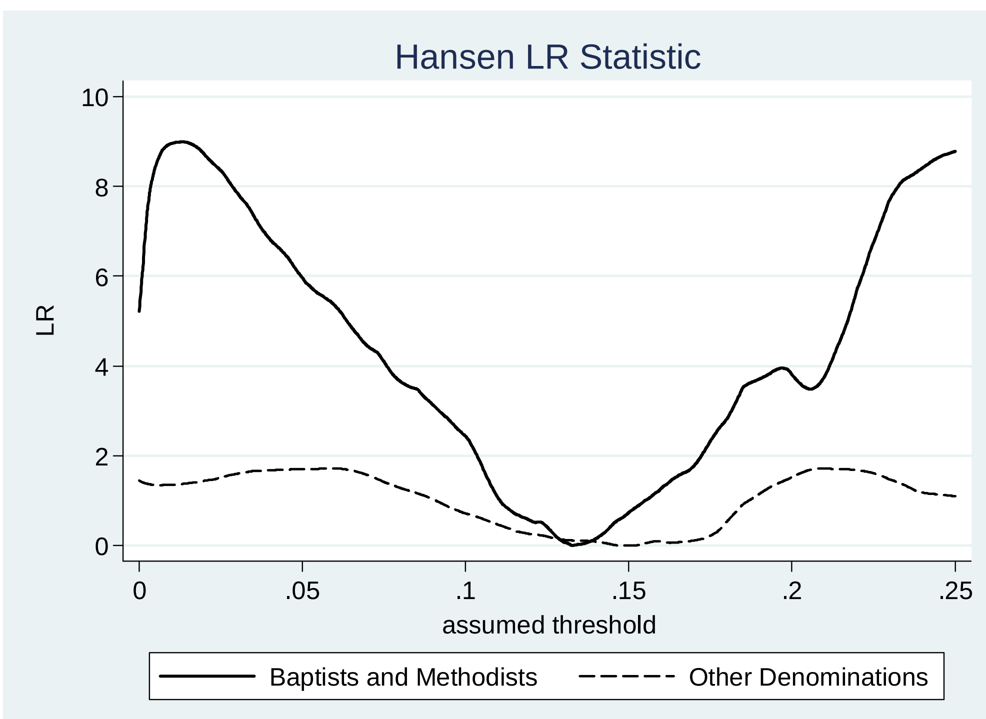


Figure A7: Hansen Test for Short Double-Difference in Black and White Population

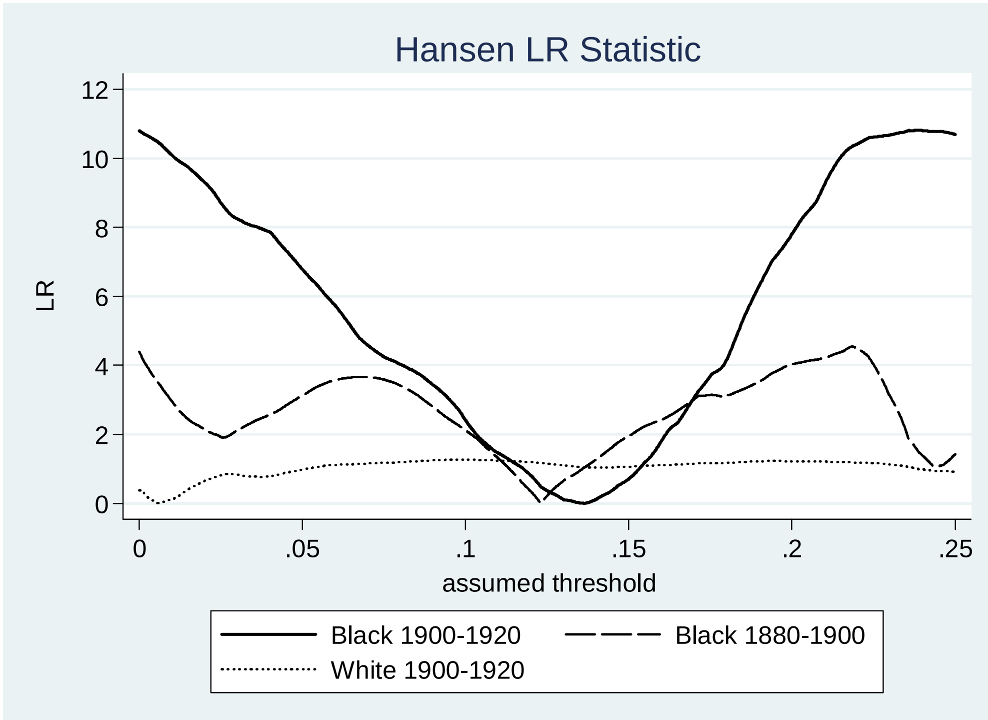


Figure A8: Hansen Test for Alternative Explanations