# Was There a Great Mississippi Flood of 1927? Estimating Impact of Southern Floods on Great Migration

Ryan Anderson
raanderson@g.ucla.edu, ECON 244, Spring 2023, Term Paper

June 16, 2023

## 1 Introduction

The emigration of southern Black Americans for northern cities in the period beginning around 1910 greatly influenced national demographics in the 20th century. The Great Migration, as this phenomenon is better known, has unsurprisingly been the subject of much study by both demographers and economists.

Some recent studies have sought to identify determinants of emigration away from Southern communities. Black, Sanders, Taylor, and Taylor (2015) tracked the impact of proximity to railroad stops with northern destinations and found powerful selection into migration via this effect. Hornbeck and Naidu (2014) investigated migration behavior and investment decisions in the wake of the Great Mississippi Flood of 1927, finding that worse-flooded counties saw much higher Black out-migration in the following years and thereby also saw greater investment into agricultural capital goods to replace the lost labor.

This paper revisits the setting of the 1927 Great Flood, employing alternate statistical methods to examine the impact of the flood on out-migration. In particular, we seek to recenter Hornbeck & Naidu's results within the framework on the impact of natural disasters on migration expounded by researchers like Leah Platt Boustan, while also demonstrating the efficacy of cluster modeling for checking the robustness of causal (or near-causal) claims.

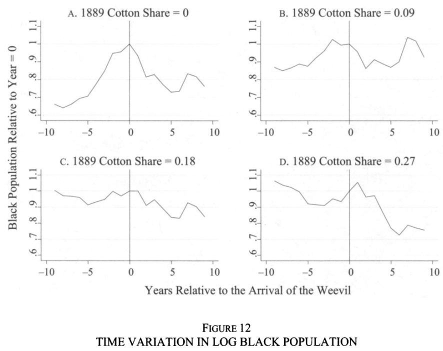
## 2 Literature Review & Theoretical Notes

### 2.1 Migration & Natural Disasters

The Great Migration saw the departure of millions of Black Americans away from the South starting around 1910 in favor of not only northern cities like Chicago, St. Louis, and Detroit, but also growing western centers like Los Angeles and Oakland. From 1914 to

1943, no less than 40% of each southern-born birth cohort had emigrated by the time they turned 40 (Black et al. (2015)).

The decision of southern Blacks to emigrate was complexly layered, and economic historians have sorted various explanations put forth into push and pull factors. One ready source of explanations comes from natural phenomena such as crop failures, pest infestations, or massive weather events, whose exogeneity might be presumed to lead to easily identifiable causal estimates. Lange, Olmstead, and Rhode (2009) track the spread of the boll weevil, "America's most destructive agricultural pest," the impacts of whose arrival on local agrarian economies two historians once compared to that of emancipation. Lange et al. find that the advance of the weevil from the Carolinas down to Texas coincided with major declines in both Black and non-Black populations, an effect more pronounced in counties with higher intensity cotton agriculture.



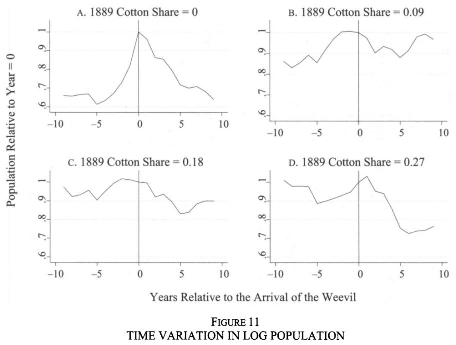
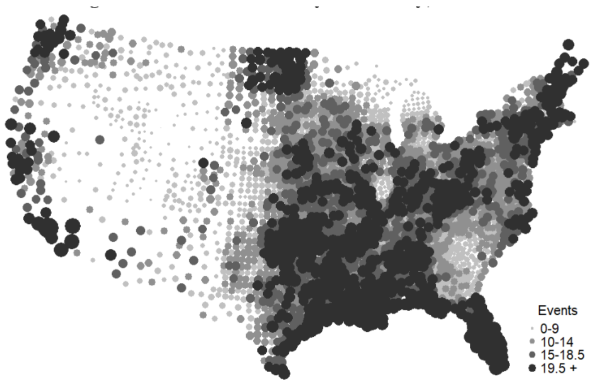




Figure 1: Black (L) and total (R) migration trends before and after arrival of boll weevil, indexed by cotton production intensity, taken from Lange et al. (2009)

Hornbeck and Naidu (2014) investigate the impact of an important disaster, the Great Mississippi Flood of 1927, on Black out-migration. The Great Flood remains the worst river-borne flooding in US history, with over 500 dead due to its effects and an estimated $17 billion in 2023 dollars of damages. Hundreds of thousands were displaced, many of them Black agricultural workers, who spent months in relief camps administered by the Red Cross. Hornbeck & Naidu describe horrific conditions of life in the camps for the displaced peoples, including whippings by National Guardsmen and an incident in which the mayor of a town in Louisiana shot a Black man for refusing to work. Substantial Black out-migration followed: one trade journal in 1927 noted that "in certain sections of the lower Delta above the Arkansas and Yazoo where a crop could not be made this year two-thirds to four-fifths of the families have moved away".

Hornbeck & Naidu employ a differences-in-differences method to estimate the impact of flooded status on changes in Black population. They find that percent of county area flooded led to strong positive shocks in out-migration, and moreover that this effect was concentrated among Black residents – whites were no more likely to emigrate from flooded counties than non-flooded counties. Overall, flooded status increased the proportion of Black residents leaving the South by 7 points, a massive effect.

More generally, a set of papers by Leah Platt Boustan and colleagues has attempted to estimate effects of different natural disasters on migration in the US over the 20th century. Boustan, Kahn, and Rhode (2012) estimated odds ratios of residency in state economic areas affected by floods, tornadoes, hurricanes, and earthquakes in two periods, 1920-1930 and 1935-40. They find that while areas affected by tornadoes lost population after the fact, areas affected by floods tended to *gain* population. This discrepancy they attribute to the difference in government responses – Congressional legislation passed in 1928 paid for expanding levee systems in the Mississippi basin, and further appropriations in the late 1930s after failures in the Ohio and Mississippi basins led to the construction of storage and runoff facilities. They hypothesize that this infrastructure investment may have enhanced not only the value of the land in affected areas but also land values in adjoining upland areas. By contrast, there was no real way to forecast or insure against the risk of tornado destruction in the period, so the migration response couldn't be affected by government investment.




Notes: This map plots disaster counts within each county for the whole period 1930–2010. The marker size is increasing in number of events, while color represents quartiles of disaster counts.

Figure 2: Frequency of natural disasters by county, taken from Boustan et al. (2020)

Boustan et al. (2020) updated the team's previous work by expanding their database to cover the period from 1920 to 2010 and including winter storms, forest fires, and other disasters. With this bevy of additional data, the authors confirmed their 2012 finding that in-migration *increases* in areas after experiencing floods – the effect is in fact more powerful earlier in the period surveyed. This effect is also robust at either the $0.05$ or $0.01$ level, depending on the controls used in the regression. The authors reprise their hypothesis from the earlier paper relating to government investment. They also point to a pattern of frequent, low severity, and predictable disasters being less likely to lead to outmigration.

The tension between the results for migration found in Boustan – low and positive – and Hornbeck & Naidu – high and negative – suggest a resolution may be found by approaching the same data with different statistical techniques.

### 2.2 Revisiting Flooding in the Mississippi River Valley

### 2.3 Hypothesis Testing via Mixture Models

In this section we briefly summarize the statistical tools used in the below analysis. We start by noting that diff-in-diff identification as Hornbeck & Naidu perform is tantamount to a form of classification, or of grouping observations into two distinct subsets marked by the treatment. Put another way, a full statistical model for Black migration decisions in the Mississippi delta might have a parameter that differed based on flooded status in the 1927 flood.

We can identify this kind of model as a mixture model. The density for a mixture model takes the form

$$p(y|\theta) = \sum_{k=1}^{K} \pi_k p_k(y),$$

where $p_k$ are our mixture components and $\pi_k$ are our mixture weights, which satisfy $0 \leq \pi_k \leq 1$ and $\sum_{k=1}^{K} \pi_k = 1$. This model says that the probability of observing $y$ given parameter $\theta$ is a convex combination of different probability densities $p_k$, depending on how each $y_i$ is weighted among the densities.

It might be useful to reconceptualize this as a hierarchical model with a hidden variable $z$ which controls group membership of the $y_i$. Let us impose a prior on $z, p(z = k|\theta) = \pi_k$ and obtain a likelihood function $p(y|z = k, \theta) = p_k(y) = p(y|\theta_k)$. Then we can marginalize out $z$ to obtain

$$p(y|\theta) = \sum_{k=1}^{K} p(z = k|\theta)p(y|z = k, \theta) = \sum_{k=1}^{K} \pi_k p(y|\theta_k)$$

In this setting, the probability distribution any $y_i$ follows is given by its hidden group membership $z_i$. So we can encode something like flooded status/non-flooded status by modeling our outcome $y_i$ as change in Black population and letting $\theta_1$ be the likelihood of Black out-migration given flooded status and $\theta_2$ be the likelihood of Black out-migration given non-flooded status. We could then specify a probabilistic model to try and estimate:

$$p(y|\theta) = \pi_1 p(y|\theta_1) + \pi_2 p(y|\theta_2),$$

where the $\pi_k$ now refer to priors on flooded status for each county. Of course, these priors on flooded status are themselves data-driven, which is slightly anti-Bayesian, but makes sense in a context of observing first the floods, then the migration responses.

It turns out that one recent fruitful avenue of research has considered how to perform statistical tests on a class of hypotheses indexed by the number of groups in a mixture model. That is, we can test $H_0 : k = 2$ vs $H_1 : k > 2$ and obtain $p$-values to determine whether we can reject or fail to reject the null hypothesis at a desired error rate. In a series of papers from 2009-2012, Jiahua Chen and Pengfei Li developed a novel and simple test based only on the use of the expectation-maximization (EM) algorithm to test the order of the mixture. Chen and Li (2009) constructs a test statistic as follows.

Given $p(y|\theta) = \sum_{k=1}^{K} \pi_k p_k(y)$, let $l_k(\pi) = \sum \log(p(y|\theta))$ be its log-likelihood function. When any of the variances of the component densities goes to $0$, this likelihood function diverges. We therefore need to penalize the function in such cases to have it converge and use it as a loss function in the optimization problem. Consider

$$\ell_k(\pi) = l_k(\pi) + \sum \rho_k(\sigma_k^2),$$

where $\rho_k$ is a smooth function of the variances of the components such that $\rho_k \to -\infty$ whenever $\sigma_k \to 0$. Let $\hat{\pi}_0$ be the maximum value of $\ell_k(\pi)$ under the null hypothesis – by an earlier result, this is a consistent estimator for the true mixing parameter $\pi$.

In order to get a test statistic, all we need to do is compute another maximum of the $\ell_k$ function under the alternate hypothesis – call this $\hat{\pi}_A$. Then a penalized likelihood ratio test statistic along the lines of $2(\ell_k(\hat{\pi}_A) - \ell_k(\hat{\pi}_0))$ would give us our desired result. Due to technical reasons having to do with the Fisher information of the mixing parameter, this doesn't work so simply, so we instead apply the EM algorithm to get a test statistic which is asymptotically close to the true test. Indeed, the theorem proven in Chen and Li (2009) provides that for an asymptotic test statistic $EM_k^N$, the test converges in distribution to a $\chi^2_{2m_0}$ distribution, where $m_0$ is the order of the mixture model under the null hypothesis.

Chen, Li, and Fu (2012) provides more detail on the construction of the test statistic, useful penalty functions to consider, and proofs for consistency. More to the point, the authors also introduced an R package which given a dataset can run the EM-asymptotic test to obtain results on the order of the mixture model used to fit the data.

## 3 Data & Results

Our data is taken from Hornbeck and Naidu (2014), graciously accessible via the AER website. The main variables for our use will be decadal county-level counts of Black and total population, as well as a dummy for flooded status which will only be used for comparison's sake.

We begin by reconstructing Fig 3, Panel A of Hornbeck and Naidu (2014), showing that the difference in change in county-level Black population by flooded status, relative to 1920. We can reproduce the chart, showing that indeed flooded counties saw a very large decline in Black population relative to non-flooded counties.

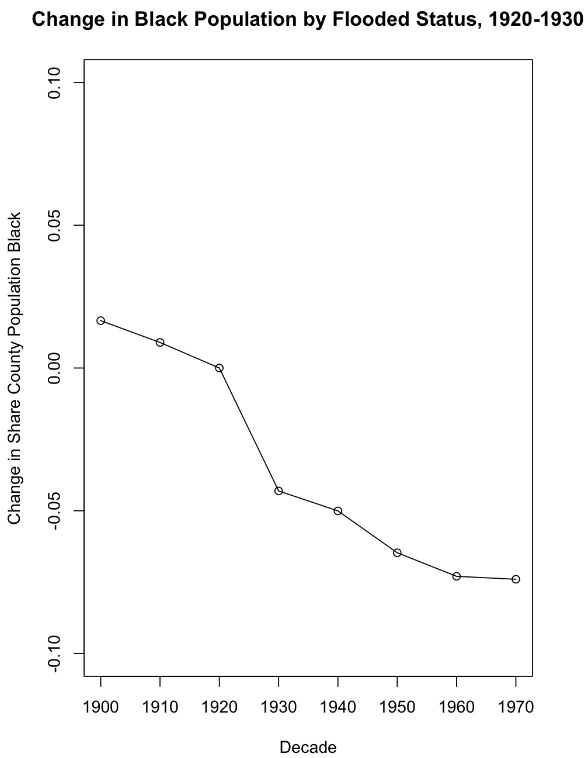



Figure 3: Change in Black Share by Flooded Status. Calculations by the author using data from Hornbeck and Naidu (2014)

We now apply our mixture model test on a set of hypotheses indexed by the number $k$, which corresponds to the count of groupings of counties as regards change in Black population 1920-1930. Somewhat surprisingly, the Chen-Li EM test provides significant evidence that there are only two relevant groups, but leaves open the possibility that more than two groups may exist. The $p$-value for the test $H_0 : k = 1$ vs $H_1 : k = 2$ is $4.09x10^{-51}$, whereas the same figure for the test $H_0 : k = 2$ vs $H_1 : k > 2$ is only $0.009$, and by the time we test 3 groups, the $p$-value falls out of significance. The below table presents the MLEs for $\pi_k$, the mixing weights, and then the means and variances of the proposed component distributions, which are normally distributed.

| **Hypotheses on *k*** | **Hypothesis** | ***π_k*** | **mean** | **variance** | **p-value** |
|---|---|---|---|---|---|
| ***H_0*: 1 vs 2** | H_0 | 1.000 | 0.012 | 0.301 | 4.09E-51 |
| | H_A | (0.095, 0.905) | (1.175, -0.098) | (1.802, 0.012) | |
| ***H_0*: 2 vs 4** | H_0 | (0.095, 0.905) | (1.175, -0.098) | (1.802, 0.012) | 9.00E-03 |
| | H_A | (0.501, 0.429, 0.02, 0.049) | (-0.074, -0.131, 3.713, 0.995) | (0.003, 0.029, 2.563, 0.126) | |
| ***H_0*: 3 vs 6** | H_0 | (0.931, 0.059, 0.01) | (-0.1, 1.005, 4.681) | (0.012, 0.069, 0.058) | 1.90E-01 |
| | H_A | (0.471, 0.46, 0.029, 0.029, 0.005, 0.005) | (-0.068, -0.134, 1.101, 0.907, 4.681, 4.681) | (0.004, 0.019, 0.054, 0.039, 0.041, 0.033) | |

Figure 4: Results from the mixture model test for three hypotheses

# 4 Discussion

The results from our mixture model test suggest that the weight of the evidence favors a two-group arrangement of counties in the Mississippi Delta as concerns changes in Black population from 1920-1930. Moreover, we do find a rather large negative effect for one of the groups in the two-group model – the average decline there was estimated to be nearly $10\%$.

The results of this test are in no way dispositive, but the fact that the model was able to recover a two-group situation without knowledge of flooded status suggests that there was something in reality discriminating among counties that led to heterogeneity in Black outmigration decisions. This may moreover suggest that the results from the Boustan literature on responses to natural disasters do not consider differential impact by race. Hornbeck and Naidu (2014) found that the impact on non-Black out-migration was essentially zero – the dynamic between non-Black non-movement and low Black population share may account for the small positive estimates found in Boustan's century-long survey.

That said, the model did find a significant result in the four-group model, with two equally sized groups having out-migration and two smaller groups having essentially zero migration. This may suggest another valid grouping which breaks along a category different to flooded status.

More work could be done to compare classification labels against Red Cross camp placements, which were not necessarily exogenously determined, but rather determined via a political process wherein local agents interfaced with both the Red Cross and the federal government (Herbert Hoover made his name providing aid after the Great Flood!).

## References


Black, D. A., Sanders, S. G., Taylor, E. J., & Taylor, L. J. (2015, February). The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South. *American Economic Review*, *105*(2), 477–503. Retrieved 2023-05-15, from https://pubs.aeaweb.org/doi/10.1257/aer.20120642 doi: 10.1257/aer.20120642

Boustan, L. P., Kahn, M. E., & Rhode, P. W. (2012, May). Moving to Higher Ground: Migration Response to Natural Disasters in the Early Twentieth Century. *American Economic Review*, *102*(3), 238–244. Retrieved 2023-06-15, from https://pubs.aeaweb.org/doi/10.1257/aer.102.3.238 doi: 10.1257/aer.102.3.238

Boustan, L. P., Kahn, M. E., Rhode, P. W., & Yanguas, M. L. (2020). The Effect of Natural Disasters on Economic Activity in US Counties: A Century of Data.

Chen, J., & Li, P. (2009, October). Hypothesis test for normal mixture models: The EM approach. *The Annals of Statistics*, *37*(5A). Retrieved 2023-06-16, from http://arxiv.org/abs/0908.3428 (arXiv:0908.3428 [math, stat]) doi: 10.1214/08-AOS651

Chen, J., Li, P., & Fu, Y. (2012, September). Inference on the Order of a Normal Mixture. *Journal of the American Statistical Association*, *107*(499), 1096–1105. Retrieved 2023-06-16, from https://www.tandfonline.com/doi/full/10.1080/01621459.2012.695668 doi: 10.1080/01621459.2012.695668

Hornbeck, R., & Naidu, S. (2014, March). When the Levee Breaks: Black Migration and Economic Development in the American South. *American Economic Review*, *104*(3), 963–990. Retrieved 2023-04-17, from https://pubs.aeaweb.org/doi/10.1257/aer.104.3.963 doi: 10.1257/aer.104.3.963

Lange, F., Olmstead, A. L., & Rhode, P. W. (2009). The impact of the boll weevil, 1892–1932. *The Journal of Economic History*, *69*(3), 685–718. Retrieved 2023-06-15, from http://www.jstor.org/stable/40263940