"Jim Crow and Black Economic Progress After Slavery"
Lukas Althoff and Hugo Reichardt
ReadMe last updated: July 7, 2024

This document describes the replication archive for the above paper. The archive contains the programs required to replicate the results in the paper. Datasets that can be shared in the archive are included. Instructions for how to download datasets that cannot be shared are also included. **Users should read this document in its entirety to successfully run the replication archive.** Questions regarding this archive may be directed to lalthoff@stanford.edu.

Contents:

1. COMPUTATIONAL REQUIREMENTS
2. DOWNLOADING OF DATASETS NOT INCLUDED IN THE ARCHIVE
3. BUILDING DATASETS FOR ANALYSIS & REPLICATING THE ANALYSIS
4. DESCRIPTION OF SOURCES FOR DATA INCLUDED IN THE ARCHIVE

---

1. Computational requirements

Software Requirements: We conducted analysis using Stata/MP 17.0-18.5.
Computational Requirements: We run code on a personal computer (128 GB), the NBER server, and an anonymous credit bureau's server.
Time Requirements: Takes approximately 48 hours to run.
Package installation: Any STATA packages that might needed are installed by 0_MASTER.do.

---

2. Downloading of datasets not included in the archive

Unzip the data folder contained in the archive. Download additional datasets as described below.

ICPSR Data (must be downloaded and stored by replicator)

United States Historical Election Returns, 1824-1968 (ICPSR 1)

Availability: This data is not provided in the replication package as redistribution to non-ICPSR members is not permitted. Users should download the data using the link below and unzip and store the folder as: data/voter_participation/additional_states/spss

Citation: Inter-university Consortium for Political and Social Research. United States Historical Election Returns, 1824-1968. Inter-university Consortium for Political and Social Research [distributor], 1999-04-26. https://doi.org/10.3886/ICPSR00001.v3

United States Agriculture Data, 1840 - 2012 (ICPSR 35206)

Availability: This data is not provided in the replication package as redistribution to non-ICPSR members is not permitted. Users should download the data using the link below and unzip and store the folder as: data/aggregate_data /county/agriculture/source_data

Citation: Haines, Michael, Fishback, Price, and Rhode, Paul. United States Agriculture Data, 1840 - 2012. Inter-university Consortium for Political and Social Research [distributor], 2018-08-20. https://doi.org/10.3886/ICPSR35206.v4

IPUMS USA Full-Count Data (must be downloaded and stored by replicator)

United States Full-Count Census Data, 1850-1940

Availability: Redistribution of full-count IPUMS USA census data is not permitted. There are two IPUMS extracts from the full-count census that cannot be provided in the replication package. These extracts must be obtained by

first registering an account with IPUMS USA at usa.ipums.org, and then creating extracts using the variables and samples identified below:

1. **usa_00043.dta**

   Select the following sample: USA FULL COUNT, 1850-1870 and 1940

   Select the following variables: YEAR, SAMPLE, SERIAL, HHWT, STATEFIP, COUNTYICP, COUNTYNHG, URBAN, GQ, FARM, OWNERSHP, OWNERSHPD, VALUEH, PERNUM, PERWT, RELATE, RELATED, SEX, AGE, MARST, RACE, RACED, BPL, BPLD, MBPL, MBPLD, SCHOOL, EDUC, EDUCD, LIT, LABFORCE, CLASSWKR, CLASSWKRD, OCC1950, IND1950, INCWAGE, INCNONWG, REALPROP, PERSPROP, OCCSCORE, SEI, PRESGL, ERSCOR50, MIGPLAC5, VERSIONHIST, MIGCOUNTY, HISTID, FBPL, FBPLD, SPLOC, MOMLOC, POPLOC

2. **usa_00044.dta**

   Select the following sample: USA FULL COUNT, 1880-1930

   Select the following variables: YEAR, SAMPLE, SERIAL, HHWT, STATEFIP, COUNTYICP, COUNTYNHG, URBAN, GQ, FARM, OWNERSHP, OWNERSHPD, VALUEH, PERNUM, PERWT, RELATE, RELATED, SEX, AGE, MARST, RACE, RACED, BPL, BPLD, MBPL, MBPLD, SCHOOL, LIT, LABFORCE, CLASSWKR, CLASSWKRD, OCC1950, IND1950, OCCSCORE, SEI, PRESGL, VERSIONHIST, HISTID, FBPL, FBPLD, SPLOC, MOMLOC, POPLOC

Select the .dta format for the extract. Save the extracts and place in data/raw/ipums.

Citation: Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek, "IPUMS USA: Version 11.0 [dataset]," Minneapolis, Minnesota: IPUMS, 2021. https://doi.org/10.18128/D010.V11.0 (accessed June 2021)

---

3. **Building datasets for analysis and replicating the analysis**

To build datasets used for main analysis and to replicate the main results, first open code/0_MASTER.do in STATA and change the home directory path to the replication directory on your drive.

- 0a_MASTER.do: Sets up the environment and install required packages
- 1a_ipums.do: Reads in the IPUMS data and saves the main dataset as well as smaller random samples from it
- 1b_linking.do: Cleans the dataset and stores the linked individual-level census datasets for Black Americans
- 1b_linking_white.do: Cleans the dataset and stores the linked individual-level census datasets for white Americans
- 1c_otherdata.do: Creates various other data files that are used in the analysis
- 1d_jimcrow_laws.do: Builds the Jim Crow database
- 2a_analysis.do: Conducts the descriptive analysis (excluding causal analysis)
- 2b_analysis_causal.do: Conducts the causal analysis (excluding RDD)
- 2c_borderpair_rd.do: Conducts the causal analysis (RDD only)
- 2d_migration.do: Analyzes migration patterns
- 2e_empirical_bayes.do: Conducts analysis using Empirical Bayes shrinkage
- 3_linking_stats.do: Prepares the linking statistics for the table
- 3_texdoc_1.do: Creates tables on the Free-Enslaved gap
- 3_texdoc_1_iv.do: Creates tables on the IV estimates of the Free-Enslaved gap
- 3_texdoc_1_plcbo1875.do: Creates tables for placebo regressions
- 3_texdoc_1blg.do: Creates tables at the census block group level
- 3_texdoc_1tct.do: Creates tables at the census tract level
- 3_texdoc_2.do: Creates tables that compares the economic status of free and enslaved around the Civil War

- 3_texdoc_2_1940.do: Creates tables that compares the economic status of free and enslaved in 1940
- 3_texdoc_linkinnov.do: Creates tables that compare the linking rates of individuals and families
- 4_nber_namebased.do: Conducts the name-based analysis on the NBER server (cannot be executed by user without access to the NBER server)

---

4. Description of sources and data included in the archive

For a detailed description of all sources and data included in the Archive, please see section B of the paper's online appendix.