Online Appendix for

"Wealth, Slave Ownership, and Fighting for the Confederacy: An Empirical Study of the American Civil War"

Andrew B. Hall Connor Huff Shiro Kuriwaki

January 24, 2019

Intended for online publication only.

Contents

A Data Construction Procedure

This section describes the specific steps we took to combine datasets and constructor our main variables of interest. Almost all steps were implemented in the R programming language.

A.1 Dataset Overview

The article deals with four individual-level, full name-containing datasets:

	Units	Smallest Geographic Information
1850 Census	All Free Citizens	County
The 1850 Slave Schedule	Slaveholders and slaves	County
Confederate State Rosters	Confederate soldier - entry	State
1832 Winner's List	Lottery winner	County of origin

Several features of the data are worth further highlighting. Only the Census comes with a set of unique numeric identifiers. We refer to the *serial ID* as IPUMS' variable `serialp`, which is constructed to identify unique households. The availability of a identifier makes the Census the preferred dataset upon which to base others.

On the other hand, the confederate state rosters pose a challenge to measurement because each record in this dataset is an entry in the roster. Thus, the same soldier may appear twice on the list if at different times he served in different units, or in the same unit but disconnected windows of time. There is no perfect way of distinguish name duplicates due to different people holding the same name with the same person entering under different units. As we detail in the following section, this challenge both makes the '$\frac{M}{N}$' weighting meaningful but also introduces other concerns.

A.2 Procedure

In our subsequent description, we use terminology from SQL or the SQL-based R package dplyr, and define them as follows:

- "Table": Any rectangular dataset with records (rows) and variables (columns).
- "Key": A set of variables that together uniquely defines each record in the table.
- "Duplicates/Is Unique": Duplicates and unique rows with regard to, unless otherwise noted, the table's key.
- "Group A by K": To collapse rows of A with the same key (K) values and summarize them by a statistic, such that the resulting table is unique in terms of its key.
- "Summarize Y by f": To apply a function f such as `mean` or `max` to a variable during grouping (collapsing), in order to define a new variable Y.
- "Subset A to B": To define a new table B by a subset of the rows of A based on some criteria.

- "Bind A and B": To stack the rows of A and B together. Row-bind.

- "Left join from A to B on K": For each row of A, find matching rows in B, where "matching" means that keys K of both records are all the same. Subsequently append those records as new variables to the corresponding A records. Note that this will never drop any rows from A. Duplicate rows of B will lead to duplicating the corresponding rows of A, so often we want B to be unique. Thus in practice this is similar to `merge m:1` in Stata.

A.2.1 Pre-processing

We pre-process the raw census as follows, and use the resulting table throughout.

1. Remove any spaces and apostrophes from the first name. Do the same for last name.

2. Drop entries where names contain the characters `*` or `?`. These characters were due to illegible census forms.

3. Drop entires where either the first name or last name is only two characters or less.

4. Drop entries with a birth year earlier than 1750, to remove outliers.

A.2.2 Combine Census and Slave Schedule

(1) Group the raw slave schedule, originally keyed on slaves, by their slave owners' full name and county. Summarize *total slaves* as the number of slave entries.

(2) Further group the slave count table by using soundex names rather than recorded names. Summarize *total slaves* again as the mean of previous slave totals.

(3) Subset the census to household heads, where these are defined by the relative codes { 101, 201, 501, 502, 601, 602 }.

(4) Left join from (3) to (2) on soundex full name, county, and state.

(5) Group the unjoined slave schedule rows of (4) by first initial (instead of first soundex name), last soundex name, county, and state. Summarize *total slaves* again as the mean of previous slave totals.

(6) Left join from the unpopulated census rows of (4) to (5) on the keys of (5). Assign *total slaves = 0* to all un-populated rows.

(7) Bind the unjoined rows of (4) and (6) itself.

(8) Group (7) by serial ID, and summarize the following household-level variables:

    - *total household slaves* by the maximum of the number of the *slave count*

- *total* $\frac{1}{N}$ *slaves* by the maximum of the member's slave count *weighted by* 1 over the number of duplicates in (3), which we shall refer to as N[1].
- $\frac{M}{N}$ *slaves* by the same idea as $\frac{1}{N}$ but where weights are further multiplied by the number of duplicates M in either (2) or (5).
- *at least one slave* by an indicator of whether *total household slaves* > 0.
- $\frac{1}{N}$ *at least one slave* as the sum of the member's $\frac{1}{N}$ weighted slave count, capped at 1.
- $\frac{M}{N}$ *at least one slave* as equivalent to $\frac{1}{N}$ *at least one slave* but with using the $\frac{M}{N}$ weights.

We end up with a slave schedule table keyed on serial ID and the above six types of variables.

A.2.3 Combine Census and Confederate Rosters

(1) Bind all confederate rosters, and generate soundex first names and last names.

(2) Group (1) by soundex full name and state. All *in roster* entries for each row is 1.

(3) Subset the census to men.

(4) Left join from (3) to (2) on soundex full name and state.

(5) Group the unjoined roster rows of (4) to by first initial (instead of soundex first name), soundex last name, and state. Summarize *in roster* again as 1.

(6) Left join from the unpopulated census rows of (4) to (5) on the keys of (5). Assign *in roster = 0* to all unpopulated rows.

(7) Bind the joined rows of (4) and (6) itself.

(8) Group (7) by serial ID, and summarize the following household-level variables:

  - *total confederate count* by the sum of individuals who have *in roster = 1*.
  - $\frac{1}{N}$ *confederate count* by the sum of individuals who have *in roster = 1*, but each weighted by 1 over the number of duplicates in (3) corresponding to that name.
  - $\frac{M}{N}$ *confederate count* by the same idea as $\frac{1}{N}$ but where the weights are further multiplied by the number of duplicates M in either (2) or (5).
  - *at least one confederate soldier* by an indicator of whether *total confederate count* > 0.
  - $\frac{1}{N}$ *at least one confederate soldier* by 1 minus the product of 1 minus the $\frac{1}{N}$ weights for each row of (7)[2].

---

[1] We generally refer to the duplicates of the census as N and the duplicates of the other datasets as M.

[2] For example, consider a serial ID containing three males with various assigned weights (they would be different because their names are different). If the three weights are $\left(\frac{1}{1}, \frac{1}{1}, \frac{1}{1}\right)$, then this

- $\frac{M}{N}$ *at least one confederate soldier* as equivalent to $\frac{1}{N}$ *at least one slave*, but with using the $\frac{M}{N}$ weights.

We end up with a confederate table keyed on serial ID and the above six types of variables.

A.2.4 Combine Census and the Lottery Winner's List

(1) To create the subset of lottery entrants, subset the 1850 census to individuals with the criteria defined in Bleakley and Ferrie (2016). Essentially this finds fathers with at least one co-residing, 1829 - 1832 Georgia-born child.

(2) Subset the census to all household members with a household head in (1).

(3) Group (1) by serial ID, randomly choosing an entrant in the rare case that there are multiple entrants in a single household. Keep track of the last name and first name duplicates among these household representatives, refer to these as N.

(4) For every record in (3) (which is keyed on household and not necessarily on name), compute the `aline` phonetic string distance between the every record in the winner's list. Specifically, first compute the string distance between the first names (not soundexed). Then compute the string distance between the last names (also not soundexed). Declare the pair a match if both distances are less than 0.05 (the distance metric ranges from 0 to 1). Keep track of the number of winner's list matches each census record collects, refer to these as M.

(5) Summarize (4) the following household-level variables:

  - *treat* $\frac{1}{N}$ by whether or not the entrant name has at least one match, divided by N the number of name duplicates in (3).
  - *treat* $\frac{M}{N}$ by whether or not the entrant name has at least one match, divided by N and multiplied by M the number of matches the pair has accrued.
  - *treat* by an indicator of whether or not $\frac{1}{N}$ is 1. This is essentially assigning a 1 to names which accrued at least one match but erring on the side of limiting false positives for those winning names in which multiple potential entrants exist.

(6) Left join from the census subset in (2) to (5) on serial ID.

This leaves us with an individual-level table where treatment is assigned at the household level.

summary will give us $1 - (0 \times 0 \times 0) = 1$. If the weights were $\left(\frac{2}{3}, \frac{1}{2}, \frac{1}{4}\right)$, then this summary will give us $1 - \left(\frac{1}{3} \times \frac{1}{2} \times \frac{3}{4}\right) = \frac{7}{8}$. If the three weights were $\left(\frac{1}{1}, \frac{1}{2}, 0\right)$, then this summary will give us $1 - \left(0 \times \frac{1}{2} \times \frac{1}{1}\right) = 1$.

A.2.5 Putting the observational dataset together

(1) Group the pre-processed census by serial ID, summarizing socio-demographic variables such as the total real estate property value.

(2) Left join from (1) to the final output of section A.2.2 on serial ID.

(3) Left join from (2) to the final output of section A.2.3 on serial ID. The order between (2) and this is irrelevant.

A.2.6 Putting the experimental dataset together

(1) Group the final output of section A.2.4 by serial ID, summarizing socio-demographic variables such as the total real estate property value.

(2) Prepare a new slave schedule and confederate roster dataset starting from the final output of section A.2.4 as the census data, but otherwise following the exact same steps as in sections A.2.2 and A.2.3. This will generate a table keyed on serial ID among the lottery entrant's descendants.

(3) Combine the three datasets in the same way as section A.2.5

A.3 Sources of Merge Error and Implications

The task of merging together the 1850 census, the roster of Confederate soldiers, and the roster of slaveowners is not a perfect process. In addition to cleaning the raw data in a number of ways, we try to reconcile slight differences in spelling between names by using the soundex encoding of first and last names. Where possible, we also use geographic information to improve our merges; nonetheless, we know that merge errors remain. It is therefore important to consider how errors in this process could affect the conclusions we draw.

As a first problem, suppose that some names are simply easier to merge than others. Hard-to-merge names in the 1850 Census will therefore be less likely to be found in the Confederate roster *and* in the slaveowners' roster. This type of error could induce a false correlation between slave ownership and fighting, because it produces faulty observations where we mark people as both non-slaveowners and non-soldiers. To try to address this problem, we re-estimate our results with the addition of last-name fixed effects. Here comparisons are only made among individuals with the same last name, which should help control for the ease of merging different names. Tables C3 and C4 present the results. The results on number of soldiers are quite similar; the results on the fraction of eligible men who fight also remain positive and meaningful, though they are smaller in magnitude than before.

A second problem is that some names, like John Smith, are very common. This is not a serious problem for matching to the slave roster, because we also know the county each person resides in, but for the Confederate roster we know only names and states, for the most part. Having the state helps to some degree, but there are still a significant number of duplicate matches. Imagine, as a hypothetical, that we had 10 John Smiths in the census, 5 of whom are slaveowners, and 5 John Smiths in the Confederate roster. Without more information that we do not have, we have no way of knowing if 5 of the John Smiths are slave-owning soldiers, or if 5 John Smiths are soldiers and non-slaveowners while the other 5 John Smiths are non-soldiers and slaveowners, or anything in between. To address this problem, we re-estimate our results only for the households in the 1850 census for which the household head has a unique name within each state. Among this set of households, we can be almost certain that, if we find them in the slave roster or in the Confederate army roster, it is the same household. Tables C5 and C6 present the results. We continue to find similar results; across all specifications, slave ownership continues to be strngly predictive of fighting in the confederate army.

B Additional Descriptive Statistics

Figure B1 shows the the geographic distribution of slaveowning households as of 1850, while Figure B2 shows the geographic distribution of Confederate army membership during the Civil War. Data for what became the state of West Viriginia is not shown.

Figure B1 – Proportion of Slaveowning Households in 1850, by County

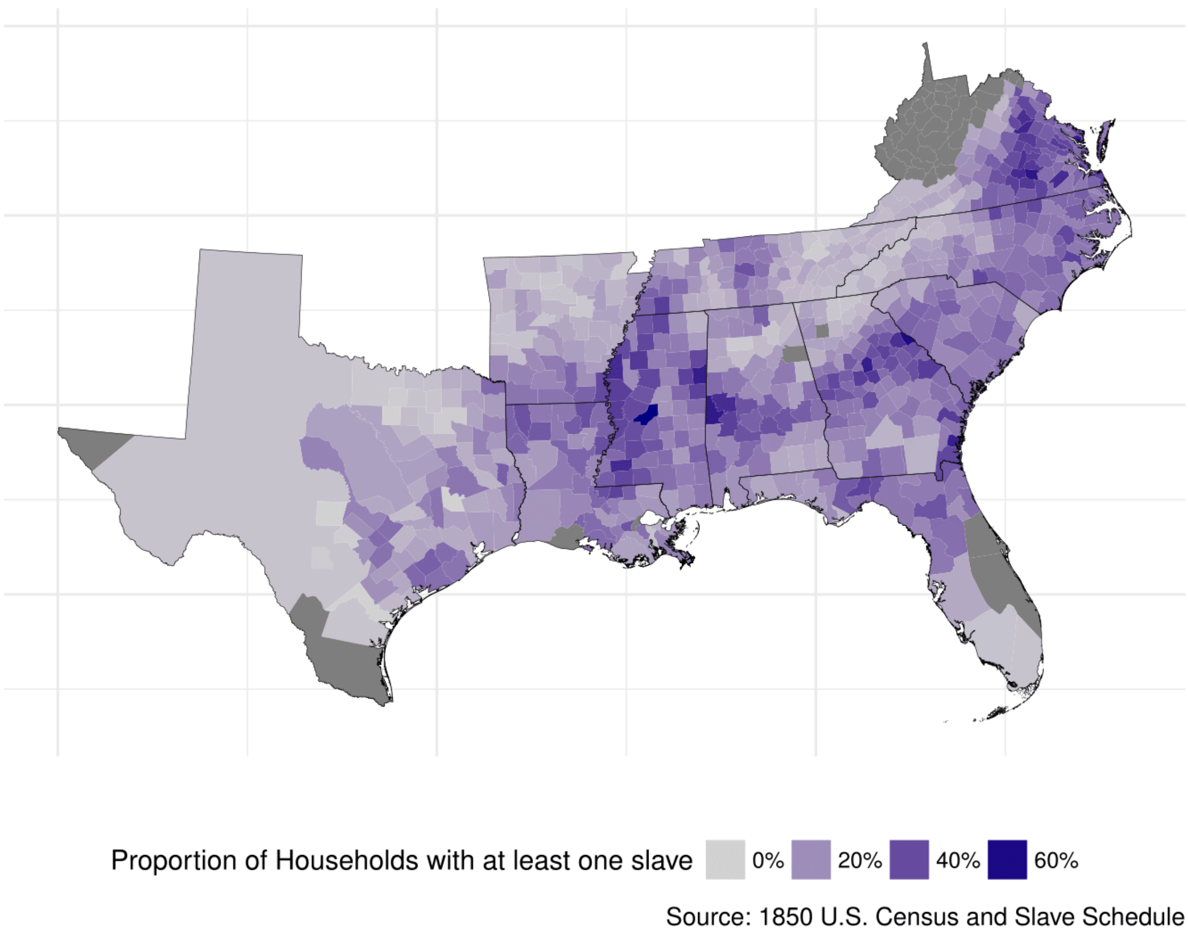
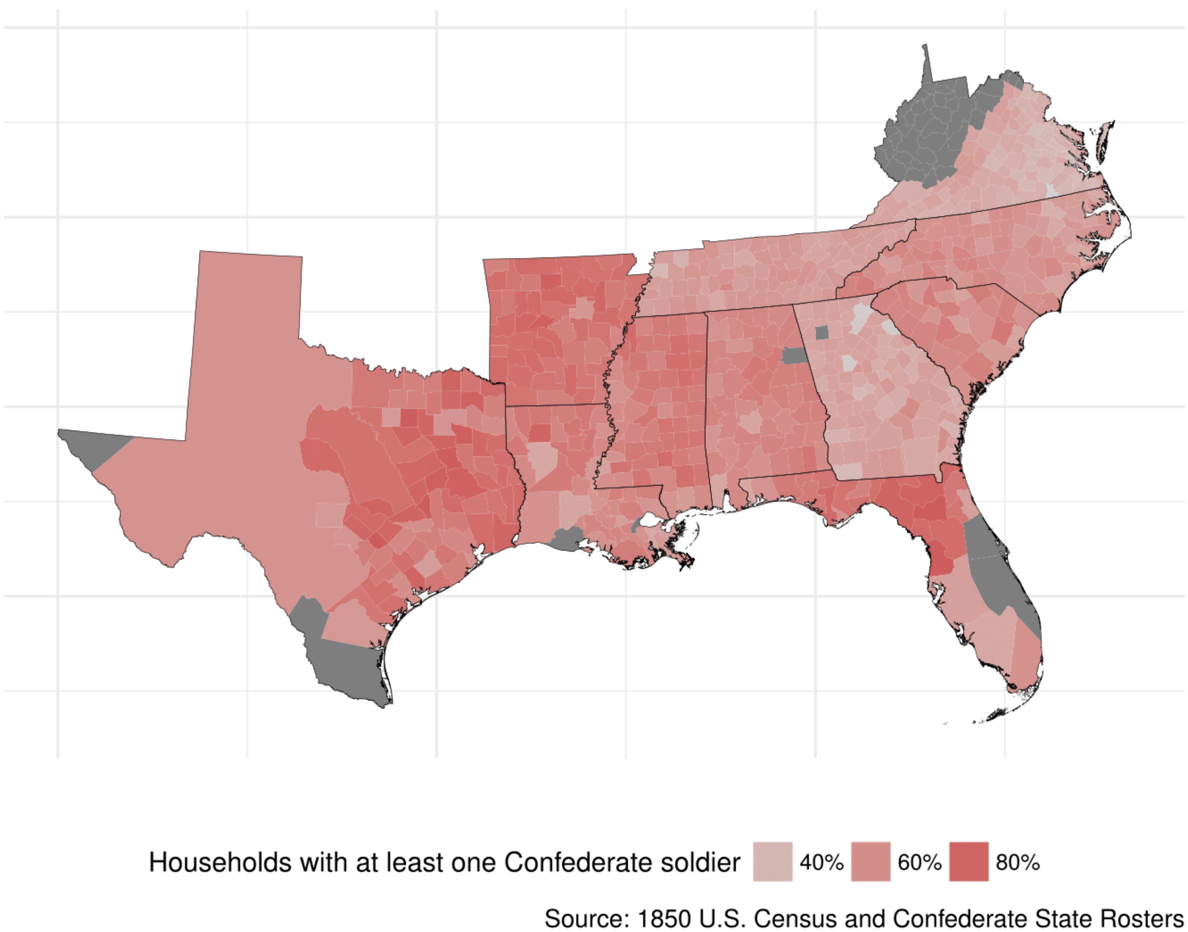


Figure B2 – Proportion of Households with at least one Member in the Confederate Army, by County



C Additional Observational Analyses

C.1 Multivariate Regressions

For a more formal treatment of our observational analyses, in the following tables we estimate the associations between slave and property ownership on fighting in the Confederate Army by multivariate linear regression.

In Table C1 we present four specifications. The first three columns show controlled difference in means: they include, respectively, a dummy variable for slave ownership, a dummy variable for property ownership, and both dummy variables in one regression. The estimates show how households that own slaves send, on average, 0.12 - 0.32 more household members to the Confederate Army, conditional on other measures of wealth. In the fourth column we examine the linear association between the degree of wealth: we add linear specifications for number of slaves, scaled as 10s of slaves, and the value of property, measured as 1,000s of 1850 dollars. These specifications are also interacted with the indicators, so that the function does not reflect the jump from non-ownership to ownership.

Both slopes are small relative to the coefficients on their dummy variable counterparts. In particular, the associated increase in fighting from a marginal increase in 1000s of dollars of real estate property value is not distinguishable from zero. The increase in army membership for slave and property owners seems to be mainly related to the jump from non-ownership to ownership.

As we have stressed, these are observational relationships. There are many possible reasons slaveowners and property owners fought in the Confederate Army at higher rates. One possible explanation is that these patterns are merely the result of wealthier households having more sons. In Table C2, we re-estimate the same specifications from Table C1, but by setting as the outcome the fraction of eligible men in a given household who fought in the Confederate Army. We still find that a higher fraction, roughly 5 percentage points more, of eligible men fought in the Confederate Army from households that owned slaves. However, the coefficient on property wealth drops, suggesting that the large coefficients on household property wealth in Table C1 may have been biased upward by a household size confounder.

C.2 Robustness to Alternative Specifications

We next ask if the positive correlations of slave ownership and wealth ownership are robust to alternative specifications and potential consequences of merge error.

First, Tables C3 and C4 re-estimate the regressions in Tables C1 and Table C2, respectively, with fixed effects and clustered errors. We apply two types of fixed effects: one for each 1850 county of residence and another for each soundex-encoded last name in the population. County fixed effects control away cross-geographic confounding in the relationship between wealth and fighting by examining individual wealth differences within a county. Last name fixed effects control away confounding variables that vary across ethnic and socioeconomic status as proxied by lineage (last name) by examining differences in wealth within a lineage. We then allow for errors within each group to be correlated by clustering standard errors by the same grouping[3].

Second, Tables C5 and C6 re-estimate the regressions in Tables C1 and Table C2, respectively, on a subset of the population that is unaffected by certain types of merge error outlined in Section A.3. We only examine people in the Census whose (soundex-encoded) full name is unique within their state, computing all measures of household-level slave ownership and fighting with this subset. In this way, we avoid the consequences of certain names being differentially duplicated during the match.

In all four tables, we see our most important finding continues to hold. That is, a household's slaveownership in 1850 is positively correlated with their estimated level of fighting in the 1860s.

[3] We do this in light of the possibility that our "treatment" of interest, slave-wealth, may exert an effect at the respective group-level. We do not cluster our standard errors in our experimental estimates concerning hte lottery because in that case our treatment variable is assigned at the individual level.

Table C1 – Predictor of Fighting in the Confederate Army. Examines the full population of white, male Confederate citizens linked to the 1850 Census. Slaveowners and property owners fought in the Confederate army at higher rates than those without slaves or without property.

	Outcome: Number of Soldiers in Household			
Owns Slaves	0.320		0.136	0.121
	(0.005)		(0.005)	(0.005)
Owns Property		0.586	0.551	0.550
		(0.004)	(0.004)	(0.004)
# Slaves (in 10s) · Owns Slaves				0.017
				(0.003)
Property Value (in 1000s) · Owns Property				0.0000
				(0.0004)

Each regression includes 746,506 households. Robust standard errors in parentheses. In column 4, number of slaves measured in 10s. Property value in thousands of 1850 dollars.

Table C2 – Predictors of the Fraction of Household Men who Fought in the Confederate Army. Estimates the same regressions as Table C1 with the fraction of men who fought as the dependent variable.

	Outcome: Fraction of Men in Household Who Fought			
Owns Slaves	0.056		0.053	0.051
	(0.001)		(0.001)	(0.001)
Owns Property		0.021	0.008	0.011
		(0.001)	(0.001)	(0.001)
# Slaves (in 10s) · Owns Slaves				0.006
				(0.001)
Property Value (in 1000s) · Owns Property				-0.002
				(0.000)

Each regreesion includes 746,506 households. Robust standard errors in parentheses. In column 4, number of slaves measured in 10s. Property value in thousands of 1850 dollars.

Table C3 – Number of Men in Households who Fought in the Confederate Army, with County and Last Name Fixed Effects. Examines the full population of households in Confederate states linked to the 1850 Census. Slaveowners and property owners fought in the Confederate army at higher rates than those without slaves or without property.

	Outcome: Number of Soldiers in Household							
Owns Slaves	0.368	0.253			0.208	0.063	0.190	0.048
	(0.012)	(0.010)			(0.015)	(0.010)	(0.016)	(0.010)
Owns Property			0.537	0.578	0.484	0.562	0.475	0.558
			(0.009)	(0.008)	(0.010)	(0.008)	(0.010)	(0.008)
# Slaves (in 10s) · Owns Slaves							0.015	0.014
							(0.004)	(0.003)
Prop. Value · Owns Property							0.004	0.001
							(0.001)	(0.001)
County Fixed Effects	✓		✓		✓		✓	
Last Name Fixed Effects		✓		✓		✓		✓

Each regression includes 746,506 households. Clustered robust standard errors, where clusters are groupings used in fixed effects, in parentheses. In column 4, number of slaves measured in 10s. Property value in thousands of 1850 dollars.

Table C4 – Fraction of Household Men who Fought in the Confederate Army, with County and Last Name Fixed Effects. Repeats regressions in Table C3 but with fraction of men as a dependent variable.

	Outcome: Fraction of Eligible Men in Household Who Fought							
Owns Slaves	0.066	0.027			0.068	0.023	0.067	0.021
	(0.004)	(0.003)			(0.005)	(0.003)	(0.005)	(0.003)
Owns Property			0.010	0.017	-0.007	0.012	-0.006	0.014
			(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
# Slaves (in 10s) · Owns Slaves							0.003	0.004
							(0.001)	(0.001)
Prop. Value · Owns Property							-0.001	-0.001
							(0.000)	(0.000)
County Fixed Effects	✓		✓		✓		✓	
Last Name Fixed Effects		✓		✓		✓		✓

Each regression includes 701,900 households. Clustered robust standard errors, where clusters are groupings used in fixed effects, in parentheses. In column 4, number of slaves measured in 10s. Property value in thousands of 1850 dollars.

Table C5 – Number of Men in Households who Fought in the Confederate Army; Unique Names Only. Examines the subset of free citizens in the 1850 Census whose full names were unique within their state of residence.

	Outcome: Number of Soldiers in Household			
Owns Slaves	0.292		0.123	0.122
	(0.004)		(0.005)	(0.005)
Owns Property		0.372	0.338	0.338
		(0.003)	(0.004)	(0.004)
# Slaves (in 10s) · Owns Slaves				0.002
				(0.003)
Property Value · Owns Property				0.000
				(0.000)

Each regression includes 547,747 households. Robust standard errors in parentheses. In column 4, number of slaves measured in 10s. Property value in thousands of 1850 dollars.

Table C6 – Fraction of Household Men who Fought in the Confederate Army; Unique Names Only. Repeats regressions in Table C5 but with fraction of men as a dependent variable.

	Outcome: Fraction of Men in Household Who Fought			
Owns Slaves	0.048		0.073	0.070
	(0.002)		(0.002)	(0.003)
Owns Property		-0.031	-0.050	-0.047
		(0.002)	(0.002)	(0.002)
# Slaves (in 10s) · Owns Slaves				0.006
				(0.002)
Property Value · Owns Property				-0.001
				(0.000)

Each regression includes 412,541 households. Robust standard errors in parentheses. In column 4, number of slaves measured in 10s. Property value in thousands of 1850 dollars.

D Randomization Validation

To validate the experimental analysis, we verify that our treated and control groups look similar on pre-treatment covariates. We conduct these balance tests for two main reasons—first, because we might be concerned that the lottery's randomization was compromised, somehow, and second, to address the possibility that our phonetic string-matching procedure, by which we located individuals in the 1850 census, could differentially match people in a way correlated with their potential outcomes of their future wealth and their propensity to fight in the Confederate Army.

Figure D3 checks for differences between lottery winners and losers on the three main pre-treatment covariates we can observe. Each panel plots the standardized difference-in-means for each of the three treatment variable specifications (the simple binary treatment indicator, the $\frac{1}{N}$ treatment variable, and the $\frac{M}{N}$ treatment variable).[4]

All differences are negligible relative to their variance, on the order of 0.01 to 0.1 standardized differences. For instance, in the left panel, we examine whether lottery winners are systematically older or younger than lottery losers. Across all three treatment variables, we see that lottery winners are very slightly younger than lottery losers. The unstandardized difference is 0.4 years in terms of birth year, not a major difference. It is almost identical to the difference reported in Table 1, p. 1467 of Bleakley and Ferrie (2016).[5] In the remaining panels we also examine standardized differences of the number of characters in entrant's names, the rate at which their last name starts with either 'M' or 'O', a common indicator for Celtic origin (Bleakley and Ferrie, 2016), and the number of children and step-children in the household born in 1832 or earlier. Most differences are negligible, except perhaps that from the $\frac{M}{N}$ indicator. As such, we will favor the other two specifications in our analyses.

E Additional Lottery Analysis

E.1 Alternative Specifications

This section offers additional robustness checks for the lottery findings presented in the article. Table E7 examines the subset of those whose names are unique within a state, Table E8 uses estimates computed by re-weighting duplicates in the $\frac{M}{N}$ specification, and Table E9 estimates adds 1850 household size as a control variable in all regressions. Further details are provided in each figure.

---

4 We use the standardized difference-in-means rather than the t-statistic as our test statistic of assessing covariate balance, following Imbens, Guido. W., and Donald. B. Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press, p.310-311. Given the large sample sizes, test statistics are likely to be statistically significant even if imbalances are not meaningfully large.

5 Bleakley, Hoyt, and Joseph Ferrie. 2016. "Shocking Behavior: Random Wealth in Antebellum Georgia and Human Capital Across Generations." *Quarterly Journal of Economics* 131(3): 1455–1495.

Figure D3 – Balance of Pre-Treatment Covariates in Lottery Sample. Tests for imbalances in four observed pre-treatment covariates from the 1850 Census: year of birth (leftmost panel); number of characters in household head's full name (center-left panel); whether or not the household head's last name starts with 'M' or 'O', a common indicator of Celtic origin (center-right panel); and the number of children in the household born 1832 – the year of the lottery – or earlier (rightmost panel). Points represent standardized differences between lottery winners and losers, using one of three treatment specifications. No substantive imbalances are found.

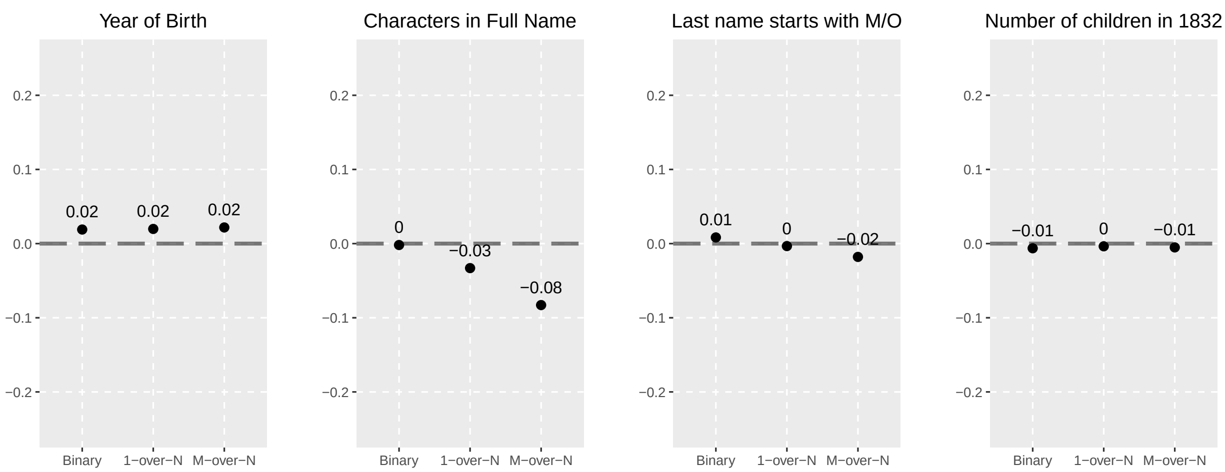


Table E7 – Effect of Winning 1832 Lottery on Household Confederate Army Membership; Unique Households Only.

	Number of Confederate Soldiers in Household		
	No FEs	Last Name FEs	First Name FEs
Won Lottery	0.29	0.14	0.25
	(0.06)	(0.05)	(0.06)
Won Lottery $\left(\frac{1}{n}\right)$	0.37	0.23	0.31
	(0.06)	(0.05)	(0.06)

	Probability at Least One Son Fights		
	No FEs	Last Name FEs	First Name FEs
Won Lottery	0.042	0.019	0.035
	(0.01)	(0.01)	(0.01)
Won Lottery $\left(\frac{1}{n}\right)$	0.050	0.029	0.041
	(0.01)	(0.01)	(0.01)

	Fraction of Sons Who Fight		
	No FEs	Last Name FEs	First Name FEs
Won Lottery	0.060	0.030	0.049
	(0.01)	(0.01)	(0.01)
Won Lottery $\left(\frac{1}{n}\right)$	0.069	0.040	0.054
	(0.01)	(0.01)	(0.01)

See experimental results in the main text for description.

Table E8 – Effect of Winning 1832 Lottery on Household Confederate Army Membership, where the dependent variable is weighted in the M-over-N specification.

	Number of Confederate Soldiers in Household		
	No FEs	Last Name FEs	First Name FEs
Won Lottery	0.35	0.20	0.34
	(0.05)	(0.05)	(0.06)
Won Lottery $\left(\frac{1}{n}\right)$	0.28	0.21	0.24
	(0.05)	(0.05)	(0.06)
	Probability at Least One Son Fights		
	No FEs	Last Name FEs	First Name FEs
Won Lottery	0.046	0.022	0.043
	(0.01)	(0.01)	(0.01)
Won Lottery $\left(\frac{1}{n}\right)$	0.030	0.023	0.021
	(0.01)	(0.01)	(0.01)
	Fraction of Sons Who Fight		
	No FEs	Last Name FEs	First Name FEs
Won Lottery	0.070	0.038	0.065
	(0.01)	(0.01)	(0.01)
Won Lottery $\left(\frac{1}{n}\right)$	0.053	0.041	0.039
	(0.01)	(0.01)	(0.01)

See experimental results in main text for description.

Table E9 – Effect of Winning 1832 Lottery on Household Confederate Army Membership, Controlling for 1850 Household Size. Same setup as the main table where each cell is a regression coefficient on winning the lottery, but all regressions now control for 1850 household size. Household size is first normalized to mean 0 for intepretability, then interacted with the (lottery) treatment variable. Only the coefficient on treatment is presented. The treatment coefficient can thus be interpeted as the treatment effect on fighting for the average-sized household, while allowing for the treatment effect to vary linearly by household size.

	Number of Confederate Soldiers in Household		
Treatment	No FEs	Last Name FEs	First Name FEs
Won Lottery	0.25	0.10	0.22
	(0.05)	(0.04)	(0.05)
Won Lottery $\left(\frac{1}{n}\right)$	0.28	0.14	0.23
	(0.05)	(0.04)	(0.05)

	Probability at Least One Son Fights		
Treatment	No FEs	Last Name FEs	First Name FEs
Won Lottery	0.040	0.019	0.035
	(0.01)	(0.01)	(0.01)
Won Lottery $\left(\frac{1}{n}\right)$	0.041	0.023	0.032
	(0.01)	(0.01)	(0.01)

	Fraction of Sons Who Fight		
Treatment	No FEs	Last Name FEs	First Name FEs
Won Lottery	0.057	0.027	0.049
	(0.01)	(0.01)	(0.01)
Won Lottery $\left(\frac{1}{n}\right)$	0.063	0.035	0.047
	(0.01)	(0.01)	(0.01)

E.2 Relationship with Civil War Battle Locations

The Atlanta Campaign invasion route appears to have run through the lands allocated in the Cherokee Land Lottery of 1832.[6] This fact raises a concern for an alternative mechanism: households who won the lottery may have moved to areas in which they perceived more danger of property destruction, which in turn prompted them to fight. Under this logic, wealthier individuals are not fighting due to their perception that the stakes of the conflict are higher due to the potential abolition of slavery. Rather, they happen to live in areas that are more likely to be threatened by the Union armies moving south.

To address this possibility we identified which counties experienced at least one Union battle during the Civil War.[7] While some county boundaries changed between 1850 and the Civil War, we use this variable to approximate the danger households may have perceived during the War.

We find that the land lottery made individuals approximately 1.7 percentage points more likely to live in a county that had a Civil War battle (Table E10), with about 8 percentage of the control group living in such counties. The small size of this effect paired with our historical knowledge about how Union generals chose their marching routes helps assuage concerns that our results are exclusively driven by this alternative mechanism. For example, as General Sherman was planning the route for his infamous march, he did so explicitly targeting wealthier areas.[8] This means that at least part of the effect we observe for the relationship between wealth and likelihood of observing a battle will be a function of the strategic choices of the Union generals, rather than individuals throughout the Southern Confederacy. The small size of this effect paired with our historical knowledge about how Union generals chose their marching routes helps assuage concerns that our results are exclusively driven by this alternative mechanism.

Table E10 – Effect of Winning the Lottery on Living in a Battle-Relevant Area.

	Lives in a County that Experienced Union Battle		
	No FEs	Last Name FEs	First Name FEs
Won Lottery	0.017	0.015	0.015
	(0.007)	(0.008)	(0.007)
Won Lottery $\left(\frac{1}{n}\right)$	0.017	0.017	0.015
	(0.007)	(0.007)	(0.007)

Each cell is a regression coefficient. The outcome variable is whether or not a 1850 household lives in a county which, during the Civil War, saw at least one Union battle. Robust standard errors in parentheses.

6 We thank a reviewer for making this point.

7 These data on the location of battles was recorded mainly by the National Park Service. We thank James Feigenbaum for sharing the tabulated data.

8 Feigenbaum, James J., James Lee, and Filippo Mezzanotti. 2017. "Capital Destruction and Economic Growth: The Effects of Sherman's March, 1850-1920." Working Paper.