Supplementary Appendix

"Let our ballots secure what our bullets have won": Union Veterans and the Making of Radical Reconstruction

Find the full Supplementary Appendix in the replication file.

Full table results for all regression analyses reported in tables and figures in the Paper and Supplementary Appendix may be found in the Supplementary Tables document in the replication file.

A Enlistment Rates Difference-in-Differences: Design and Robustness

A.1 Design

In this paper, I apply a difference-in-difference design to examine the effect of enlistment rates in the Union Army on county-level voting for Republicans and Black Suffrage. Here, the treatment, D_i is the enlistment rate in the Union Army among military-aged males in county i. The the time variable T_t is an indicator for whether county i is observed before the Civil War (0), and thus before the county experienced the treatment, or after the start of the Civil War (1), when the county was affected by enlistment.

I include county-specific intercepts μ_i and state-year intercepts γ_{st}. This deviates from the classical model in that it assumes parallel trends **within states**, rather than nationwide.

This application of the DID differs in two key ways from the classical formulation.

1. Continuous Treatment: In the research design for this paper, treatment is the enlistment rate; it can take on values between 0 and 1. This contrasts with classical difference-in-difference approaches in which treatment is binary.

2. All units treated: In this design, all counties have some enlistment. Thus, there are no units that go "untreated" in the post-treatment period. Nevertheless, the level of treatment that counties receive after the war is different.

These two differences raise questions: Do the assumptions required for identification differ? And what is the effect that is identified?[1]

A.1.1 Continuous Treatment

To understand what the parallel trends assumption means and what is identified, we need to expand the potential outcomes framework laid out above.

Rather than each unit in each time period having two potential outcomes (associated with treated or untreated), we can imagine that each unit i in time period t has a response schedule of potential outcomes: $Y(d)_{it}, d \in D$, where $Y(d)$ indicates the value Y would take if the continuous treatment D took on the value d.

We can represent this response schedule as $f_{it}(D)$: a function that tells us what the potential outcome of Y would be for unit i at time t for all possible values of D. This $f_{it}(D)$ could be non-linear and heterogeneous across i. The new effect we want to estimate is $\tau = E[f'_{i1}(D)]$: the average derivative of the response functions across all units i. As Angrist and Krueger (1999) point out, it is natural to view this average derivative of the response function as related to the conditional expectation function of Y as a function of D: $E[Y_i|D_i = d]$. It then makes it natural that we might estimate τ using least squares.

Just like the binary DID, we only observe $f_{i1}(D_i = d)$: the value of the response schedule for unit i for the actual value of treatment d that the unit takes. And because we do not observe the entire response function for unit i, $f'_{i1}(D)$ and $E[f'_{i1}(D)]$ are unknown.

Just as before, we can use a difference in differences to identify $E[f'_{i1}(D)]$, with some assumptions. The assumptions are easiest to see if we consider the average difference-in-differences in Republican voting between counties with an enlistment rate of d and counties with an enlistment rate of $d - \Delta$, were Δ is some arbitrary difference.

$$\{E[Y_{i1}(d)|D_i = d] - E[Y_{i0}(0)|D_i = d]\} - \{E[Y_{i1}(d-\Delta)|D_i = d-\Delta] - E[Y_{i0}(0)|D_i = d-\Delta]\}$$

This can be decomposed in terms of the average reponse function, as follows:

$$= E[f_{i1}(d) - f_{i1}(d-\Delta)|D_i = d] +$$

$$\{E[f_{i1}(d-\Delta) - f_{i0}(0)|D_i = d] - E[f_{i1}(d-\Delta) - f_{i0}(0)|D_i = d-\Delta]\}$$

This rearrangement gives us $E[f'_{i1}(d)]$ evaluated at d plus a bias term. This bias is the difference between the counterfactual trend units with enlistment rates of d would have had with enlistment rates of $d-\Delta$ and the actual trend units with enlistment rates of $d-\Delta$ had.

[1] I am not here attempting to justify an ecological interpretation of these results. Though I should note: conceiving of this model as an ecological regression sidesteps the issues of continuous treatments and all units being exposed to treatment, while introducing other problems.

Angrist and Krueger (1999) note that assumption we have to make, in a regression context, is that conditional on covariates X, this bias term goes to zero. For the continuous difference-in-difference used in this paper, the assumption is that conditional on unit i fixed effects and state-year, dummies X_{it}:

$$E[f_{i1}(d-\Delta) - f_{i0}(0)|X_{it}, D_i = d] = E[f_{i1}(d-\Delta) - f_{i0}(0)|X_{it}, D_i = d-\Delta]\}$$

If this parallel trend assumption is correct: the unit fixed effects net out the baseline differences in the response functions: $f_{i0}(0)|D_i = d$ and $f_{i0}(0)|D_i = d-\Delta$, $f_{i1}(0)|D_i = d$ and $f_{i1}(0)|D_i = d-\Delta$. And the state-year fixed effects net out any changes in $f_{i1}(D) - f_{i0}(0)$ that are shared by units with different enlistment rates.

Note: this assumption still does not assume linearity in the response function $f_{i1}(D) - f_{i0}(0)$. And, this assumption does not require that the response functions are homogenous across units. It only requires that, within states, the response function of the change $f_{i1}(D) - f_{i0}(0)$ is, on average, the same across different rates of enlistment $D_i = d$. Thus, just as in the case where all units have some treatment, the key difference is that we must extend the parallel trends assumption: the average causal response functions of units with different levels of are parallel with each other.

When might this be violated?

1. These assumptions would be violented if units with different levels of D have, on average, different shifts in the average response function $f_{i1}(0) - f_{i0}(0)$ in the absence of treatment: pre-treatment trends are non-parallel. This is easy to test, as we can examine the pre-treatment period to assess this difference in average response function for several periods of time in which $D = 0$ for all units. This is true in the binary diff-in-diff

2. Additionally, these assumptions would be wrong if there were a confounder that is correlated with enlistment rates d that also determined Republican voteshare in counties in the post-war period (but not before). This is a kind of time-varying confounding, and it would not be detected by assessing parallel trends pre-treatment. This, too, is a threat in a binary diff-in-diff.

3. Even if units have parallel trends over time in the absence of treatment, and there is no time-varying confounder, these assumption would be violated if units with different levels of treatment d have different causal response functions for $f_{i1}(D) - f_{i0}(0)|D_i = d, d \neq 0$ *on average*. This implies that the EFFECTS of treatment D — relative to unit baseline and shared trends over time —are heterogeneous. And that this heterogeneity is not independent of the levels of treatment. In order for this to induce bias, there must be both heterogeneous effects and confounding (the heterogenous effects are related to treatment assignment).

How does this parallel trend assumption differ from the binary case? The only key difference is that it assumes that, within states, the average causal response function of the change in Y from $t = 0$ to $t = 1$ is either: (i) identical for all counties i, in which case selection based on potential outcomes into d is not a problem; or (ii) the average causal response functions are heterogenous, but selection into d is independent of these heterogeneous effects.

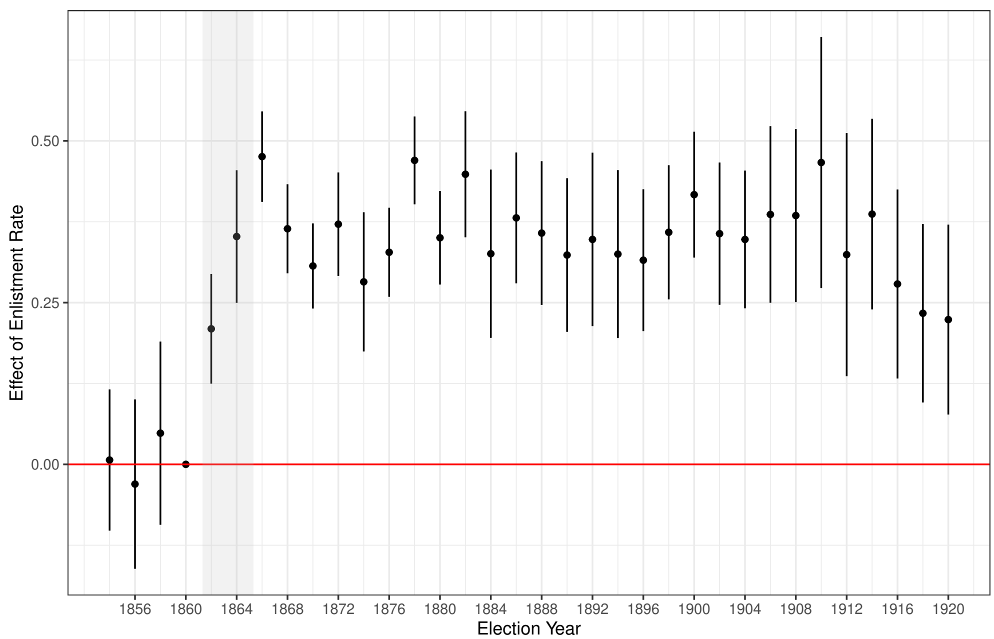
A.2 Checking Parallel Trends

I first assess pre-war parallel trends by estimating the following model, using the same data as Table 1.

$$\text{Republican Voteshare}_{\text{ie}} = \alpha_i + \alpha_e + \sum_{y=1854;\neq 1860}^{1920} \beta_y \text{Enlistment Rate}_i * \text{Year}_y + \epsilon_i + \epsilon_y \quad (4)$$

This estimates a separate slope on enlistment rate, for the difference between county Republican voteshare in each federal election between 1854 and 1920 and the 1860 federal elections. Figure A1, shows the results of these estimates. Prior to the Civil War, the over-time change in the average causal response functions of counties with different levels of enlistment is not different from 0.

Figure A1: Effect of veterans on Republican Vote-share in federal elections for each year between 1854 and 1920



This figure plots the year-specific effect of enlistment rates on Republican vote-share for federal elections with 1860 as the reference year. The model includes county and state-election fixed effects. Data from Congressional and Presidential elections across 384 counties between 1854 and 1880. Counties with election cycles in which the GOP does not contest an election cycle are dropped. Standard errors clustered by county and election year. Bars show 95 percent confidence intervals.

One might be concerned that this lacks credibility: do we really believe that counties in, say the top and bottom quartiles, of enlistment share parallel trends. In the body of the paper, I check this in Figure 1. Even when we split counties into the quartiles of enlistment in their state (reflecting the same with-in state parallel trends I specify in the model), the trends are visibly parallel across quartiles. Statistical tests of these differences confirm that the change in Republican voteshare, relative to 1860, in counties with different quartiles of enlistment are not different from 0.

A.3 Addressing Time-Varying Confounding

Another concern might be that, even if there are parallel trends for counties with different levels of enlistment before the war, there may be some other attribute of counties with high enlistment that affected Republican voteshare once the war started. To address this possibility, I first explore the predictors of enlistment in Figure A2.

As discussed in the paper, enlistment rates are significantly correlated with several economic, demographic, and political attributes of counties in 1860. These could shape how counties respond to the war. To address this possibility, I estimate a new model, including an interaction between each of these pre-war attributes and the post-war period (Table A1) or election year (Figure A3).

$$\text{Republican voteshare}_{ie} = \alpha_i + \alpha_e + \beta \text{Enlistment Rate}_i \cdot \text{Civil War}_e + \mathbf{X_i} \cdot \text{Civil War}_e + \epsilon_i + \epsilon_y \tag{5}$$

The results are substantively unchanged, even though we allow for possible time-varying confounding.

A.4 Addressing Heterogeneous Effects

Another possible source of trouble would be if the effects of enlistment varied across counties with different levels of enlistment. One way to address this problem would be to directly estimate heterogeneous effects in the enlistment DD model for counties that were otherwise similar to each other on baseline characteristics. One way to do this is to use *generalized random forests* (Athey, Tibshirani and Wager 2018). Athey, Tibshirani and Wager (2018) 's causal forests estimate the heterogeneous effects treatment W for case i. If X_i is a set of covariates and W_i is a continuous or binary treatment, generalized random forests estimate $\theta(x)$, or the effect of W on Y for a set of values of X. This is done by generating weights that define the closeness of other observations to i within the space defined by X. These weights for the closeness of case j to case i are $\alpha_j(x_i)$. The sum of $\alpha_j(x_i)$ across all j is 1. Whereas many methods generate these weights using kernels which is subject to the curse of dimensionality, generalized random forests use random forests (repeated iterations of regression trees) to assign weights. In this case, θ is the locally weighted linear partial effect of W on Y.

$$\hat{\theta}(x_i) = \frac{\sum_{j=1}^{n} \alpha_j(x_i)(W_j - (\sum \alpha_j(x_i)W_j))(Y_j - (\sum \alpha_j(x_i)Y_j))}{\sum_{j=1}^{n} \alpha_j(x_i)(W_j - (\sum \alpha_j(x_i)W_j))^2}$$

Figure A2: Predictors of County Enlistment Rates (Standardized Coefficients)

This figure plots the standardized coefficients from bivariates regressions of county enlistment rates on 32 different covariates, with state fixed effects. Data from 384 counties, dropping counties with values in top and bottom 0.5 percentiles of the respective covariate. HC1 Standard Errors. Bars show 95 percent confidence intervals.

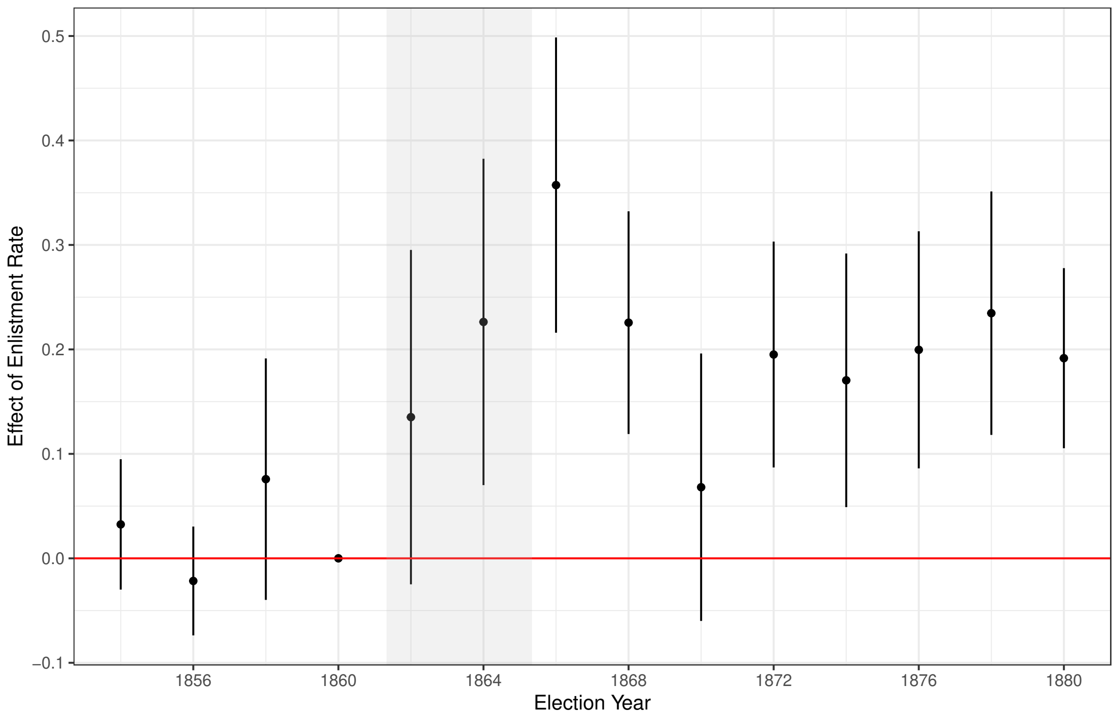
Table A1: Difference-in-Differences Estimate (With Time-Interacted Controls) of Effect of County Enlistment Rate on Republican Voteshare

	Dependent variable:		
	Republican Voteshare		
	(1)	(2)	(3)
Enlist % · Postbellum	0.175^{***}	0.192^{***}	0.177^{***}
	(0.031)	(0.043)	(0.036)
GOP no contest	included	dropped	dummy
County FE	X	X	X
State-Election FE	X	X	X
Covars · Postbellum	X	X	X
N Counties	330	257	330
Observations	6,930	5,397	6,930
R^2	0.785	0.770	0.840

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Data from Congressional and Presidential elections across 330 counties between 1854 and 1880. Standard errors clustered by county and election year. Counties with election cycles in which the GOP does not contest an election cycle either treated as a 0, the election is marked with a dummy, or the all observations for that county are dropped. The post-war indicator is interacted with 32 demographic, economic, and political covariates. Counties with missing data or extreme outliers in the covariates are dropped.

Figure A3: Effect of veterans on Republican Vote-share in federal elections for each year between 1854 and 1920, with Covariate-Time interactions



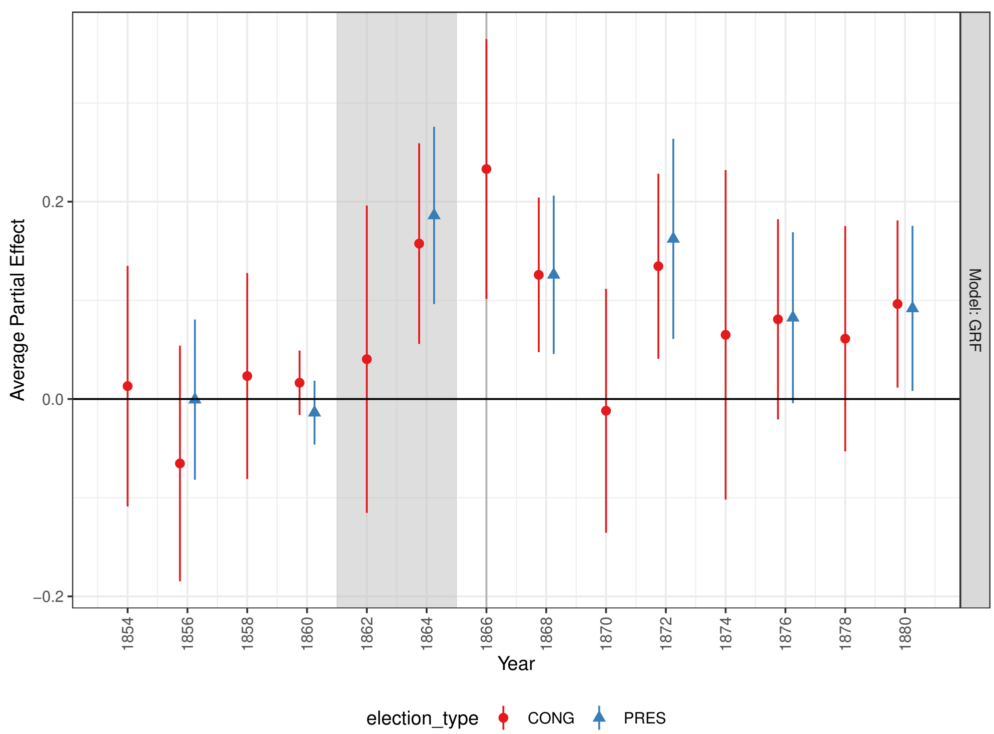
This figure plots the year-specific effect of enlistment rates on Republican vote-share for federal elections with 1860 as the reference year. The model includes county and state-election fixed effects. The year indicators, with 1860 as the reference, are interacted with 32 demographic, economic, and political covariates. Standard errors are clustered by county and year. Bars show 95 percent confidence intervals. Data from Congressional and Presidential elections across 257 counties between 1854 and 1880. Standard errors clustered by county and election year. Counties with election cycles in which the GOP does not contest an election cycle are dropped. Counties with missing data or extreme outliers in the covariates are dropped.

These individual heterogeneous effects are then combined into average partial effect, with asymptotically unbiased standard errors. Using this method, I estimate the following difference-in-difference model for each election year y between 1854 and 1880. c indicates that I have centered the variable around the state mean, to impose the state-year fixed effects.

$$\text{GOP Voteshare}_{yc} - \text{GOP Voteshare}_{1860c} = \theta(x)\text{Enlistment Rate}_{ic} + \epsilon_i \qquad (6)$$

The variables used in weighting case similarity for computing heterogeneous effects are: (a) The performance of the Republican party in every federal election (Congressional and Presidential) from 1854 to 1860. For any county in which a given party did not compete in an election, I set their vote-share as 0 and add a dummy indicating the party did not compete in this county-election or that the data is missing. (b) All non-political covariates shown above in Figure A2. (c) Indicators for the state in which the county is located. The key assumption here is that the effects of enlistment are effectively constant or independent of enlistment for cases that are similar in covariates X.

Figure A4: Average Partial Effect of Enlistment on GOP Voteshare: Diff-in-Diff



This reports the generalized random forests estimates of the average partial effect of enlistment for a difference-in-difference between Republican voteshare in each federal election between 1854 and 1880 and 1860, conditioning on political, demographic, and economic covariates. Standard errors are robust. Bars show 95 percent confidence intervals.

Figure A4 shows that the results obtained using generalized random forests are substantively the same, even though we explicitly and flexibly model heterogeneous effects of enlistment.

A.5 Robustness

The effects are unchanged when using surviving veterans as a fraction of 1860 military aged males, instead of all enlistments: Table A2.

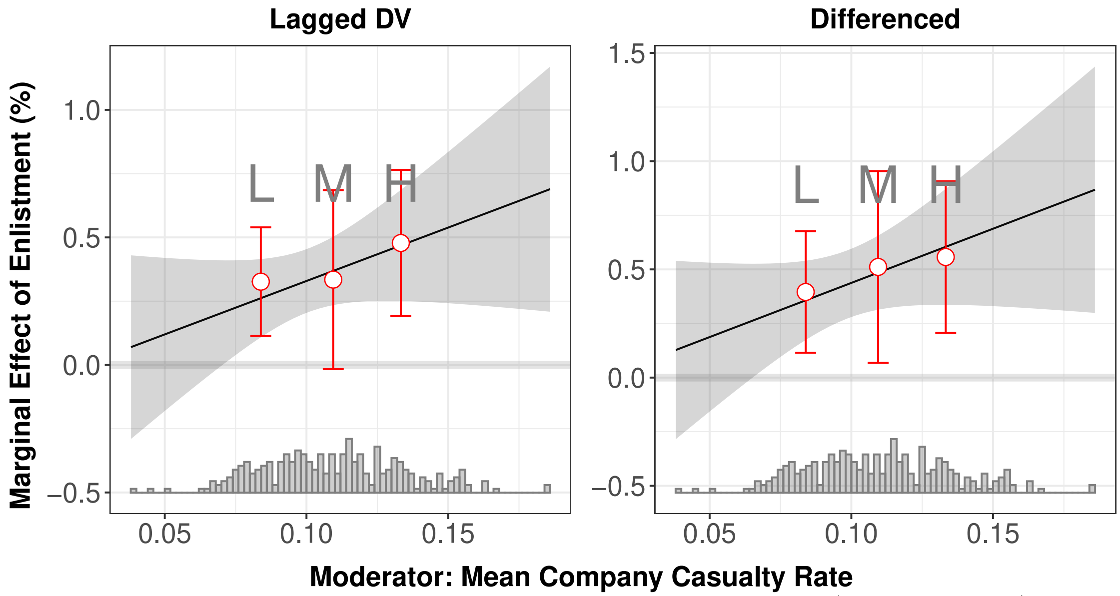
Table A2: Difference-in-Differences Estimate of Effect of County Enlistment Rate (Survivors Only) on Republican Voteshare

	Dependent variable:		
	Republican Voteshare		
	(1)	(2)	(3)
Survived % · Postbellum	0.414**	0.323***	0.322***
	(0.104)	(0.076)	(0.062)
No. Rep. Candidate			−0.585***
			(0.029)
GOP no contest	included	dropped	dummy
County FE	X	X	X
State-Election FE	X	X	X
Observations	6,153	4,473	6,153

Note: *p<0.05; **p<0.01; ***p<0.001

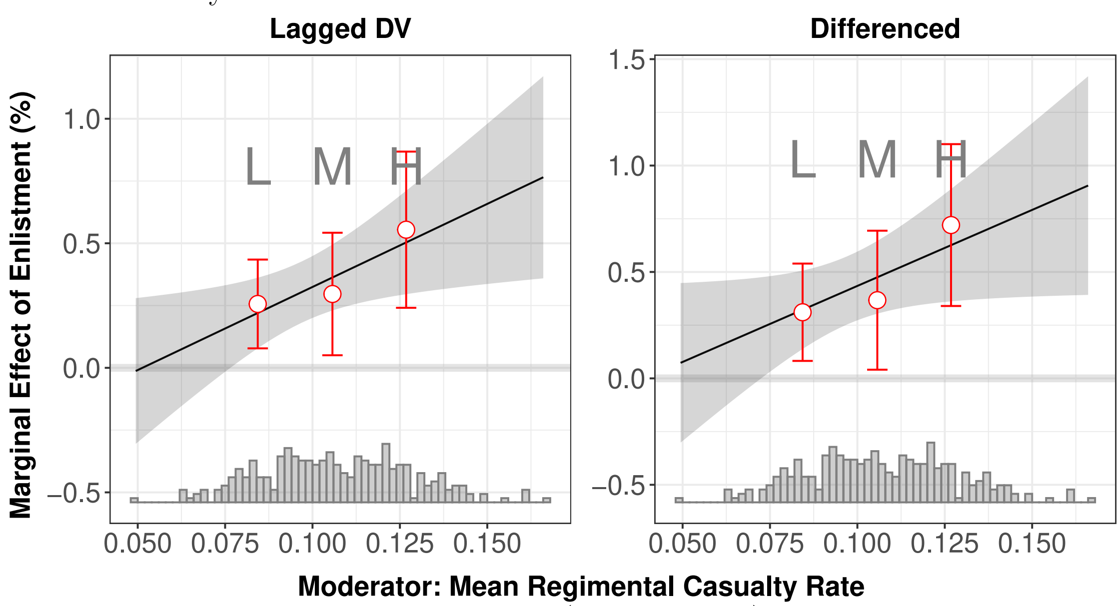
Data from Congressional and Presidential elections across 293 counties between 1854 and 1880. Standard errors clustered by county and election year. Counties with election cycles in which the GOP does not contest an election cycle either treated as a 0, the election is marked with a dummy, or the all observations for that county are dropped.

Figure A5: Marginal Effect of Enlistment Rates on Republican Voteshare Conditional on Company Casualty Rate)



This figure plots the marginal effect of County Enlistment Rate (survivors only) on Republican Voteshare in 1864-1868 elections vs 1854–1860, across Mean Company Casualty Rate, conditional on state fixed effects. Sample includes 293 counties, and excludes Indiana for which we cannot calculate mean casualty rates. Standard errors are robust (HC1).

Figure A6: Marginal Effect of Enlistment Rates on Republican Voteshare Conditional on Regimental Casualty Rate



Marginal effect of County Enlistment Rate (survivors only) on Republican Voteshare in 1864-1868 elections vs 1854–1860, across Mean Regimental Casualty Rate, conditional on state fixed effects. Sample includes 293 counties, and excludes Indiana for which we cannot calculate mean casualty rates. Standard errors are robust (HC1).

B Company Casualties in Infantry Regiments: A Natural Experiment

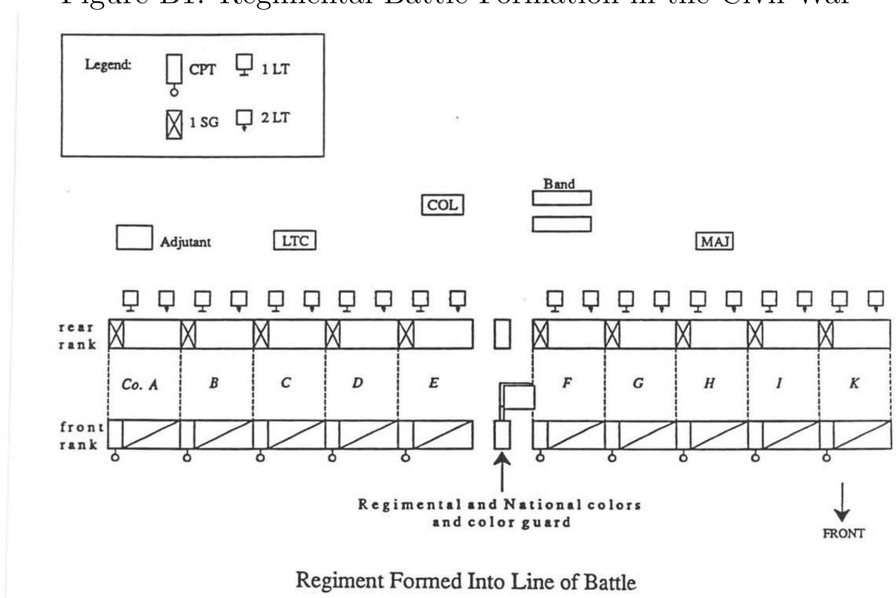
B.1 Company Exposure to Combat Fatalities

The vast majority of men serving in the Union Army served in infantry regiments. The basic unit in which infantrymen were mobilized, maneuvered, and went into combat was the regiment. While the Regular Army had independent companies of sharpshooters, volunteer regiments, which made up most of the army, operated as units. Regiments were further subdivided into ten companies of equal size, each commanded by a captain. Men in the same company were mustered in together and then trained, lived, worked, and fought directly alongside each other.

Men may have chosen when to enter military service, the branch of service, the term of their service, and even the specific regiment they joined. And Army commanders might decide based on regimental performance or perceived reliability to give certain regiments garrison duty or combat roles, or in battle, give more or less important or difficult objectives.

But, because regiments were the main unit around which military tactics of the time were designed, men in the same regiment went the same places, at the same times, and were located in the same place on the battlefield. In battle, infantry regiments typically formed up in a line of men, two deep, horizontally arranged by company, approximately 140 yards across (see Figure B1) (Hess 2015). In the chaos and smoke of battle, which companies in that line received more casualties was effectively arbitrary.

Figure B1: Regimental Battle Formation in the Civil War



This figure shows the standard battle formation of both Union and Confederate infantry regiments during the American Civil War. Ballard, Ted, and Billy Arthur, *Chancellorsville Staff Ride: Briefing Book*. Washington, DC: United States Army Center of Military History, 2002. Public Domain Image from Wikipedia

Because this logic applies only to infantry regiments, I exclude artillery and cavalry regiments from my analysis. This is because, artillery regiments deployed batteries to different places on the battlefield, and cavalry regiments, particularly early in the war, operated with detached companies.

Moreover, because the staff/headquarters company and regimental band were deployed behind the ten main companies, I exclude these from analysis. Similarly, regimental records often include "unassigned" troops, without a known company. I exclude these from analysis since we do not know which company they served with.

B.2 A Justification of As-If Random

Following Dunning (2012), I consider the information, incentives, and capabilities of soldiers, officers, and the enemy that might lead soldiers to experience more or fewer company casualties as a function of potential outcomes of partisanship. The evidence below comes almost exclusively from (Hess 2015), which is the only contemporary academic historical work devoted to Civil War infantry tactics.

B.2.1 Soldiers

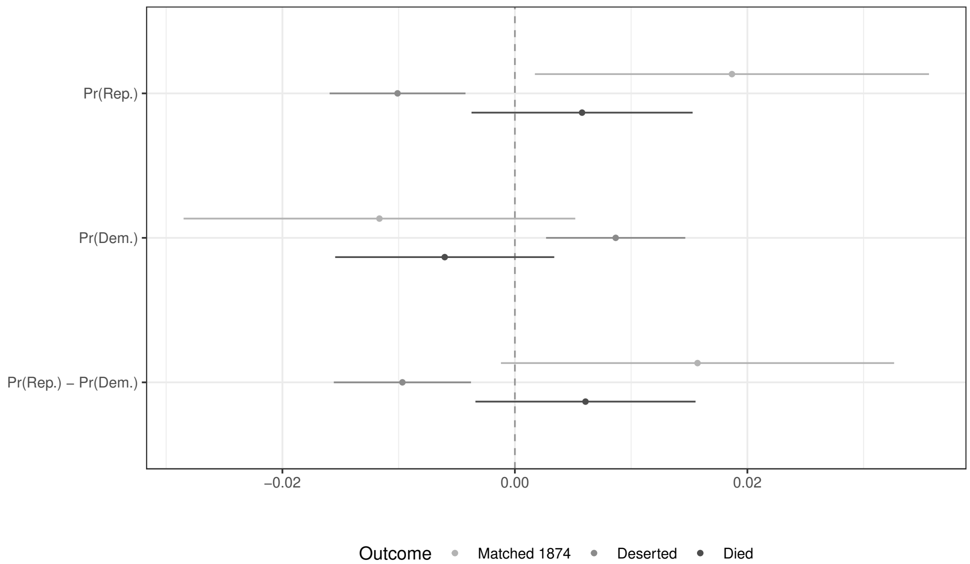
Information While soldiers could choose which company they joined when regiments were formed (Costa and Kahn 2008, 52–57), they could not have known which companies were more likely to receive casualties. (i) All companies stood together in a line across. (ii) While drill manuals specified which position companies should take in that line, soldiers entered the war unfamiliar with military drill instruction, and often the guidelines were ignored. This was true, because Army manuals called for companies to be arranged by the seniority of officers. This was irrelevant in almost all volunteer regiments, since all officers joined at the same time.

On the other hand, once in a regiment, soldiers could have known that tasks such as skirmishing (fighting in loose formation in advance of the main line) or serving in the color guard (protecting the flag bearers of the regiment) or color company (immediately to the right of the color guard) might be more dangerous. And soldiers may have known that their regiment was soon to enter battle, as well as whether some battles might be more dangerous.

Incentives In general, soldiers may have incentives to desert or use caution to avoid death. Conversely, soldiers may have wanted to seek glory or fight for the cause, leading them to take greater risks. And it is plausible that these incentives varied by partisanship. Republicans may have been more supportive of the cause and thus more motivated to take greater personal risks and fight harder, while Democrats may have been more likely to want to simply make it through the war alive. This is borne out in a raw comparison of predicted Democrats versus predicted Republicans: predicted Republicans were more likely to die, while predicted Democrats were more likely to desert (Figure B2.

Capacities Despite some information, soldiers had limited capacity to select themselves into treatment.

Figure B2: Differential Attrition by Predict Partisanship (Bivariate)



This figure shows the change in attrition, death, or desertion (related to attrition) associated with a 2 SD change in the predicted probability of Republican or Democratic partisanship in a bivariate regression model. Sample includes men serving in Indiana Regiments who were matched to the 1860 Census. Matches only include 'best' post-war links for each individual. Individuals are weighted by 1 over the number of matches. Latent partisanship is predicted probability of being Republican or Democrat based on the person's name, birth year, and birth place listed in the 1860 census. Standard errors are clustered by individual.

Before combat (i) Soldiers could work to be chosen for the color guard, which protected the regimental colors. Because the colors were essential for guiding the line of soldiers into battle and were source of morale, the enemy frequently targeted the colors. However, the members of the color guard were detached from their companies and were composed of corporals and sergeants. Excluding men of these ranks from analysis would address this form of selection.

(ii) Soldiers could not *choose* to join a company in advance based on its proximity to the colors. These companies might receive more casualties (though see below). These positions were often dictated by the company's ability to drill, which Hess notes was often a function of how skilled and motivated the captain was to drill his men. The companies best at drill would up on the flanks or in the center, near the color guard. This was to ensure that the regiment kept good formation in battle. While soldiers might attempt to shirk at drill to avoid being the color company, it seems unlikely that this would affect mortality. First, good performance at drill was also understood to be essential to for the regiment to survive under fire; even for reasons of survival, soldiers had reason to be competent at drill. Second, it is unlikely that individuals could conspire to get an entire company to be worse at drill. Third, even if they could get better at drill, it is implausible that this could be done with such precision that it would lead them to be placed on the flanks (potentially less exposure) vs. the center (more exposure). Moreover, the company directly to the left of the color guard was also likely in more danger. Yet this position was not dictated by drill performance.

(iii) Finally, soldiers could choose to desert rather than fight. This is a more serious problem of selection. If we simply were to use the actual "Exit" date from a company, soldiers who either deserted or died would determine their extent of exposure to combat casualties. And this could induce selection bias. But, I address this problem by constructing an intent-to-treat exposure to combat casualties.

During Combat Soldiers had limited capacities to alter their exposure to death during combat. Linear combat (fighting in lines) provided few opportunities for individuals or companies to take independent action. As discussed below, the whole point of having men fight in lines was to increase the effectiveness of massed rifle fire by largely unskilled marksmen under conditions of low visibility and limited means of command and control. Armies positioned officers and non-commissioned officers to keep men from lagging. And they practiced drill in order to keep their lines in good order. While men could elect not to fire, or fail in their aim due to fear, there were few opportunities to run or hide. In Hess's review of combat reports, he records very few mentions of companies operating independently at all. As discussed below, these instances were the result of battlefield idiosyncrasies.

Similarly, there were few opportunities for soldiers and companies to fight *harder*. While there are many accounts of units fighting to their last, which could induce selection bias, these accounts detail regiments, because companies simply did not operate independently. At the individual level, it could be that Republican soldiers took more personal risks. This appears to be the case, but even if this was true: (1) we don't include individuals in their own treatment variable and (2) if so, this could lead to attrition that reduces the number of Republicans observed after the war, because the most Republican men in a company might end up dying, increasing the "treatment'' for less Republican men who, because they survived, were located after the war. This would bias my estimates downward.

B.2.2 Commanding Officers

Information Officers likely knew, to some extent, the partisanship of their companies. They might infer this from where units were raised, the captains of those units, or the ethnic or occupational composition of the regiment. Or, more simply, given the vigorous political discourse within the Union Army, through conversation and debate (McPherson 1997). Based on this, officers may have been more suspicious of units that were less reliable.

Incentives There were three major considerations in how officers deployed companies in combat. **Keeping Unit Formation:** Army manuals suggested putting veteran companies on the ends and near the colors, but in an war where most regiments were of volunteers raised at the same time, this advice was rarely followed. Instead, regimental commanders debated whether to put the most well-drilled marchers on the ends of the line and in the center to help keep the regiment's line from falling apart.

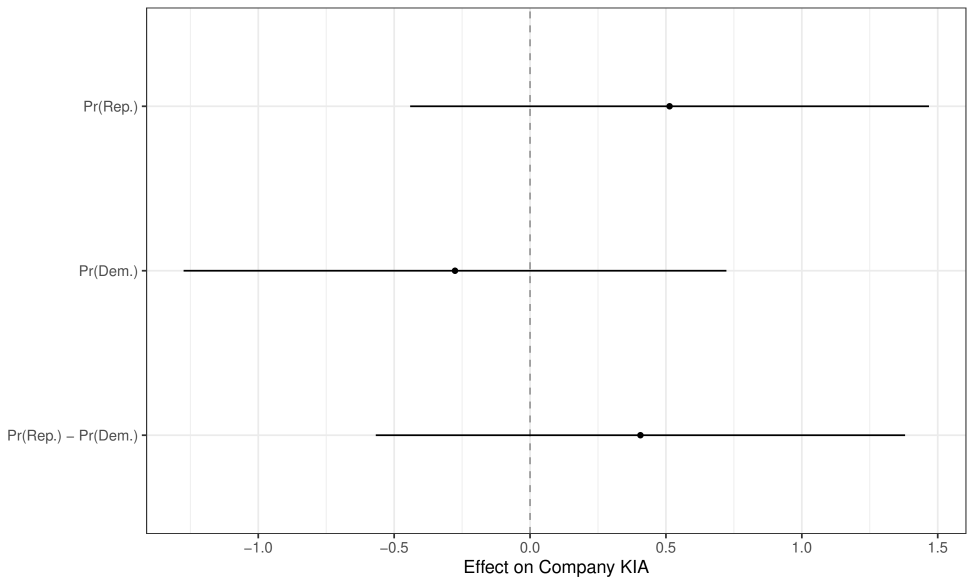
Another consideration was **terrain**. As a unit marched into combat, companies could become detached to the side or the rear of the regiment if they came upon obstacles. The affected companies moved to the rear of the rest of the regiment and retook their place in line when the obstacle was passed.

Finally, regimental tactics gave companies distinct orders based on **battlefield exigencies**. While most of the time, companies moved together in a line, deviations from this reflected emergencies. Sometimes companies were deployed as skirmishers, in loose formation, in front of a regiment. While this became more common over time, Hess notes that it was more common to use entire regiments as skirmishers for a brigade. Thus, it is unclear whether there was selection into skirmishing duty by partisanship. And it is also unclear whether skirmishers were in more or less danger, given that they were more spread out and encouraged to make use of cover. In other times, the left or right wings of regiments were used separately. This was done if the regiment were accidentally split in battle or when facing a flanking maneuver from the enemy. Again, it is hard to imagine how partisanship of companies would be used to inform how these decisions were made in the heat of battle.

If there were some selection, we'd expect companies with different baseline partisanship to have different combat casualties. In Figure B3, I report the within-regiment relationship between company casualty rates and average partisanship of men in those companies. There are no significant differences in casualty rates by company average partisanship, and the estimates are substantively small: within regiments, a 2 SD increase in mean Republican partisanship is associated with a 0.25 or smaller SD difference in casualties.

Capacity Other than choosing which companies would occupy the flanks and the center, officers could not position companies in line based on their partisanship. Fighting was done as a regiment. And when companies were detached, the reasons had to do with unexpected circumstances generated by chaos on the battlefield. One reason regiments fought together in cohesive lines was that it was difficult on a smoky battlefield, with rifle and artillery fire, to give specific directives to different companies. This is why regiments practiced drill together, in order to accomplish carefully choreographed maneuvers to be used in combat in order to fight more effectively and save their own lives in an attack.

Figure B3: Within-Regiment Relationship between mean Company Partisanship and mean Company KIAs



This figure the relationship between a 2 SD change in mean company predicted partisanship and the number of combat casualties in a company, conditioning on regimental fixed effects and weighting companies by the number of men use to calculate mean company partisanship. Results from least squares regression of mean company casualties, by company, on mean company partisanship, conditional regiment-enlistment year fixed effects. Sample includes serving in Indiana Regiments who were matched to the 1860 Census:10329 unique soldiers, in 896 companies, across 288 regiments. Companies are weighted by the number of soldiers. Standard errors HC robust.

When Hess discusses case studies of tactical errors that led to greater casualties, these occurred due to errors made by brigade or division commanders sending regiments into the wrong place, or when whole regiments found themselves in the wrong place on the battlefield.

B.2.3 Enemy

Finally, decisions made by the enemy about how to fire on regiments might create non-random company casualties.

Information Confederate units would be able to see color guards, since these were intended to be visible on the battlefield. It is also possible that they could distinguish officers.

Incentives Battle reports indicate that both armies sought to kill officers and color guards to both demoralize units and eliminate leadership to reduce the fighting effectiveness of the enemy.

Capacity While there was capacity, despite smoke, for Confederate units to see and target color bearers, and perhaps officers, this is mitigated by a few other considerations. First, officers were dispersed throughout the regiment (Captains and lieutenants were attached to each company). Second, while color bearers were targeted more, it is also the case that regiments were trained to fire against the enemy lines, not specifically at individuals. The primary reason that both armies kept using these linear tactics, despite the advent of the rifled musket, was that in order to maximize firepower, they wanted to mass rifle fire against enemy units. This was to increase the efficacy of rifle fire. Men in both armies lacked sufficient training to shoot accurately, the armies could not afford to expend munitions training them to shoot more accurately, and without smokeless powder, battlefields had limited visibility. This limited the effectiveness of individual firearms and prioritized large volumes of fire from large groups of men simultaneously.

B.2.4 Color Guards

Based on the preceding the discussion, the biggest potential source of selection effects in company-level casualties is related to color guards and color companies. Color guards, selected from corporals and sergeants of other companies, were at the center of the line. These men attracted more fire from the enemy. To their right, was the color company, often chosen for its ability to perform marching and drill well. Color guards may have been more likely to be Republican, given that it was a dangerous position. Color companies may have been more likely to be Republican, since good marching order might have been related to enthusiasm for the war.

It is impossible to reconstruct the membership of the color guard or color companies. The composition of the color guard changed and is not usually recorded on company rosters in the data available. Companies chosen to be color companies changed in regiments throughout the war: regiments often held drill competitions between companies to choose which were best.

While we cannot easily test whether color companies or selection into the color guard led to bias, a couple points suggest this is less of a problem.

(1) If members of the color guard were more likely to be Republican, then the remaining men in their companies would be exposed to more company deaths, if color guard members were more likely to die. But this would produce selection where the remainder of the company was, on average, less likely to be Republican, more likely to have more combat death exposure, and more likely to be observed after the war. This differential attrition would bias the effects I estimate downward.

(2) Companies with better marching skills might be chosen to be the color company. But better marching companies were also chosen to be on the left and right flank, far away from the color guard. And the company to the left of the color guard was not chosen based on its skills. Thus, even if companies that were more Republican on average were better motivated at drill, they could not easily determine whether they ended up in the center or on the flanks. And companies with poorer marching skills might still end up near the color guard. Moreover, the survival of a regiment in the face of the enemy depended on effective drill. Being able to complete complex choreographed maneuvers under fire often made the difference between holding off an enemy attack or being flanked and suffering great losses. Thus, efficacy in drill may have been motivated by enthusiasm for the war or enthusiasm to make it home alive.

For these reasons, I consider it implausible that soldiers could have selected into different levels of company casualties based on partisanship.

B.3 Data

In the following sections, I describe the data used in the analysis of this natural experiment and how it was collected.

B.3.1 Post-war Partisanship

This analysis is only possible because there are data on post-war partisanship for individuals who served in the war. I draw on the *People's Guides* of nine Indiana counties, published in 1874.[2] These were published by Cline and McHaffie. These guides report the history of the county and township, and include a directory of people residing there. An example of one page is located in Figure B4. While not fully exhaustive, they include a large fraction of adult men in these counties. For each person, their name, occupation, location of residence, birth place and year, date of settlement in the county, religious affiliation, and political partisanship is listed. A sample from these guides was previously collected and analyzed by (DeCanio 2007; DeCanio and Smidt 2013). Similar guides exist for 6 counties in Illinois. These are harder to digitize and contain less data on which to match individuals (no birth year), so I did not prioritize this data collection.

[2] Bartholomew, Boone, Hamilton, Henry, Hendricks, Johnson, Montgomery, Morgan, and Vermillion Counties

Figure B4: Example Page from 1874 People's Guide

COLUMBUS TOWNSHIP. 143

Apel, Chas.; carpenter and builder; Columbus. Born in Prussia 1842; settled in B. C. 1865. Rep. Lutheran.

Aikens, David; farmer; 2 m s Columbus. Born in Va. 1835; settled in B. C. 1860. Dem. Protestant.

Aikens, James; at leisure; 2 m s Columbus. Born in Va. 1828; settled in B. C. 1867. Rep. Protestant.

ABBETT, W. A.; justice of the peace and farmer; 3 m s w Columbus. Born in B. C. 1832. Indpt. Methodist.

Abbett, O. P.; farmer; 2 m s w Columbus. Born in B. C. 1834. Indpt. Methodist.

Abbett, Henry; cooper and farmer; 2 m w Columbus. Born in B. C. 1832. Dem. Christian Union.

Abbett, Washington; farmer; 2½ m w Columbus. Born in Ky. 1827.

Arwin, John S.; physician and surgeon; Columbus. Born in Tenn. 1824; settled in B. C. 1868. Dem. Protestant.

Anderson, S. F.; 4 m n Columbus. Born in B. C. 1833. Rep. United Brethren.

ARNHOLT, WM.; farmer; 4 m s e Columbus. Born in Germany 1846; settled in B. C. 1868. Dem. Lutheran.

Arnold, Thomas; farmer; 3½ m n w Columbus. Dem. Chris.

Anthony, Joseph; farmer; 3 m n Columbus. Born in Tenn. 1830; settled in B. C. 1849. Rep. Protestant.

Armet, Charles; farmer; 2½ m n w Columbus.

Akins, C. E.; collector of Singer Sewing Manufacturing Co.; Columbus. Born in Ohio 1846; settled in B. C. 1873. Rep. Christian.

Adams, D. W.; druggist; Columbus.

This figure shows an example page listing biographical entries in one of the 1874 People's Guides.

To collect the names, birth years, birth places, and partisanship of individuals listed in these guides, I did the following. (1) I downloaded all pages of the guides from archive.org. (2) I converted each page into a gray-scale image. (3) I passed each gray-scale image to Google Cloud Vision to detect the text on each page. (4) Using custom R and Python scripts, I combined letters into words, words into rows, and rows into biographical entries using the x-y coordinates of text on the page.

Then, I used regular expressions to extract the names, birth places, birth years, years of settlement, and partisanship of each individual. I then had workers on Amazon Turk correct errors in machine transcription for each biographical entry. While some minor errors likely remain, I have key biographical details for 27169 people across these 9 counties.

B.3.2 Union Soldiers from Nine Indiana Counties

To find the effects of wartime experiences on post-war partisanship, I needed a baseline sample of soldiers to track over time. While some soldiers undoubtedly left their county of residence, a comparison of soldiers from the same county seems more plausible than to include soldiers who later moved into these counties. Moreover, by considering, say, all Union Soldiers or all Union Soldiers from Indiana, I was increasing the risk of making incorrect matches.

Based on records from the Indiana Adjutant General, 21056 men enlisted in the Union Army from these nine counties. However, a large fraction of Indiana soldiers in the ACWRD lack places of residence. To link individual soldiers to these nine counties I did the following: (1) I linked each listed place of residence given by Indiana soldiers in the ACWRD to a list of all place names (county, township, city/village, post-office) reported in the 1860 and 1870 Census in Indiana. In some cases, the same place name appears in multiple counties. And for some soldiers, no place of residence is listed. (2) For soldiers whose residence uniquely matches one of the 9 counties, they were linked to that county. (3) If soldiers' places of residence matched to multiple counties in that list of 9, they were linked (if possible) to the one county to which the majority of men in his company were uniquely matched. Otherwise, they were considered as possible residents of multiple counties (4) For soldiers with no residence listed, they were assigned to one of the 9 counties if the majority of the men in their company came from that company.

Table B.3.2 shows the number of enlistees recorded in each county by the Adjutant General of Indiana and the number of individual soldiers I match to each county (weighted by 1 over the number of counties to which they match).

For these set of Union soldiers, I have the following variables: rank at enlistment, term of enlistment, type of regiment (infantry, artillery, cavalry), date of enlistment, regiment of service (and muster in/muster out dates), company of service, how this person entered the regiment (volunteer, drafted, substitute), how they left the regiment, and exposure to casualties as described in the paper.

B.3.3 Linking Soldiers to 1874 People's Guides

I link Union Soldiers to the 1874 People's Guides using a deterministic matching procedures. While Enamorado, Fifield and Imai (2019) argues that there are gains to be made by using

Table B1: Number of Union Army Servicemen by County

County	Soldiers (Adjutant General)	Soldiers (ACWRD)
Bartholomew	2813	2490
Boone	2442	3196
Hamilton	2272	2756
Hendricks	2416	2747
Henry	2549	2633
Johnson	2033	1500
Montgomery	2850	2903
Morgan	2114	2046
Vermillion	1567	1029

an automatic and probabilistic method, applying the fastLink algorithm to this data yielded poor results. The primary reason for this are the few characteristics on which to match and a substantial number of cases in which only first initials are available.

My deterministic matching procedure uses the following attributes: county of residence, first name, last name, and birth year.

For each soldier who *survived the war*, I implement these steps to identify a set of matches:

1. I identify a set of people in the 1874 People's Guides who (a) reside in the same county the soldier resided in at enlistment and (b) are reported as moving to that county before 1866 (since soldiers may have returned home as late as that). 2. I clean first names by linking reported names, which may be misspelled or abbreviated to full names, using a crosswalk created by Abramitzky et al. (2019). 3. I then generate potential matches under the following conditions:

- If the first name and last name match exactly
- If first name matches exactly and lastnames match exactly using the metaphone sound code or have a distance of less than 0.1 on a Jaro-Winkler score.
- First names match exactly on metaphone sound code or have a Jaro-Winkler distance is less than 0.1 and last name matches exactly
- For soldiers with only a first initial listed, I consider as matches those with the same first initial and an exact match on the last name.

Within these, I consider as the best matches those that are closer on all name matching metrics and have the closest year of birth.

This procedure generates matches for 4895 Union soldiers. On average, each soldier is matched to 1.51 people post-war, with a max of 14. When restricting to best matches, the average number of matches is 1.18, with a max of 12. In cases where soldiers have more than one match, I weight them in subsequent analyses with $1/m$ where m is the number of matches.

B.3.4 Linking Soldiers to the 1860 Census

In addition to linking soldiers to the 1874 People's Guides, I also link them to the 1860 US Census. I do this for three reasons. (1) Soldiers may vary in how easily they can be located, in general, in historical records. This may have to do with how consistent they were in providing biographical details, and it may reflect whether soldiers are in fact from the county I match them to using the procedure used above. Linking soldiers to the 1860 US Census provides confirmation that soldiers were in fact from the county in question and provide a pre-treatment measure of "findability".

(2) In addition to military enlistment details, this provides additional demographic characteristics on which to test balance and use as conditioning variables. (3) Most importantly, linking to the US Census permits me to generate a measure of predicted partisanship for these soldiers, even if they are not located in the 1874 People's Guides. This permits conditioning on predicted partisanship and checking for differential attrition by baseline partisanship.

I link soldiers using the FastLink algorithm, blocking on county of residence (Enamorado, Fifield and Im 2019). I generate matches between soldiers and people in the Census using on cleaned first names, last name, birth year, and the metaphone sound codes of the first and last name. I kept matches in which the posterior probability of a correct match exceeded 0.8.

In using this procedure, I match 12103 soldiers to the 1860 Census. The mean number of matches per soldier is 1.14, and the maximum number is 14. When using Census data for soldiers, I take an average of the data for each Census match, weighted by the match probability given by fastLink.

B.4 Predicting Partisanship

In order to examine imbalance on partisanship by treatment and to investigate the possibility of differential attrition by partisanship, I need to have a pre-treatment measure of partisanship for soldiers. While it is not possible to find individual partisan affiliation for soldiers in this period of time, the next best option, if imperfect, is to generate predicted partisanship based on demographic attributes of soldiers.

To do this, I use demographic data available in the 1874 to predict partisanship. Because I need to generate these predictions for soldiers who don't appear in the 1874 People's Guides and these predictions shouldn't suffer from post-treatment bias, the demographic details need to exist in both the 1874 guides and the 1860 Census. I use names, birth years, and places of birth.

I trained a machine learning algorithm using these variables to predict partisanship in 1874. I then use this model to predict the partisanship of soldiers in 1860. In what follows, I describe this process in more detail and provide validation that the predict partisanship measure meaningfully captures variation in partisanship.

B.4.1 Training Data

To generate predictions of partisanship, I use all people listed in the 1874 People's Guides who were *not matched to any soldiers*. I then created the following features corresponding to these people:

- The first name of the person as listed in the guide/census
- The last name of the person as listed in the guide/census
- The cleaned first name of the person, using the name variant/abbreviation crosswalk created by (Abramitzky et al. 2019).
- The metaphone sound encoding of the first name as listed in the guide/census
- The metaphone sound encoding o the last name as listed in the guide/census
- The birth year listed in the guide/census
- The place of birth listed in the guide/census. Here, I used state of birth in the US or country of birth (collapsing many regions in Germany/Ireland to the country).
- Partisanship, collapsed to: Republican, Democrat, Other, or None. The vast majority of people are in Republican or Democrat. The most common "other" is the Grange.

Note that the county of residence or township and occupation, although also in the census, are not included. These may be affected by war-time experience, and so I do not include them as features to predict partisanship.

B.4.2 Machine Learning Classifier

I then used the `fastText` algorithm to classify individuals' partisanship. This algorithm is used for text classification and, rather than using a bag of words to classify texts, represents words and subsets of characters in that word in a lower dimensional space, and then uses these word representations to classify documents. I used this approach, because name spellings are too numerous to include as binary-encoded features and I would be predicting with new datasets that might even contain new name variants. `fastText` permits making predictions even with names/words that have not been seen in the training data.

To train the algorithm, I split this data into 5 different training/testing groups and trained the model on each group. Each time, I used `fastText`'s auto-tune functions to select the optimal model parameters, as chosen by performance on the test group.

B.4.3 Validating Predictions

To validate the performance of these models for four sets of data. (1) I generated predictions for each model on the "test" group (not used in training, only in choosing model parameters). (2) I also check the performance of taking the average of predictions from all five models (bagging) for the people in the 1874 guides that were matched to soldiers, and not used in the training at all ("validation" group). (3) I averaged the predictions from the five models for all men of voting age in residing in Indiana counties in 1860, using data from the 1860 Census. This is the "aggregate census" group. (4) For analyses in the paper, I generate latent partisanship for soldiers by (a) averaging predicted probabilities of partisanship from all five models for people in the Census and (b) averaging these scores across all matches a soldier has in the Census, weighting by the match probability derived using fastLink. This is the "individual census" group.

Predicting Partisanship in 1874 I assess the predictive power of these classifiers in two ways.

ROC/PR Curves The ROC curve plots the trade-off between false positives rates and true positive rates, given different probability thresholds for making a binary classification. A poor classifier performs about as good as or worse than random guessing (which would have the performance of the diagonal line). The PR curve plots the trade-off between precision (the fraction of positive classifications that are true positives) and recall (the fraction of the true positives that are correctly identified by the classifier). Here, we assess the performance of the curve against the performance of the "no-skill'' classifier of guessing every case to be a match. The more the area under the PR or AUC curve, and the more these curves are higher than the "no-skill'' curves, the greater the predictive power of the classifier.

Predicted Probability vs. Actual Probability I also plot the (smoothed) probability of being a Republican/Democrat across predicted probabilities of being a Republican/Democrat. This should be nearly perfectly correlated, if the classifier performs well.

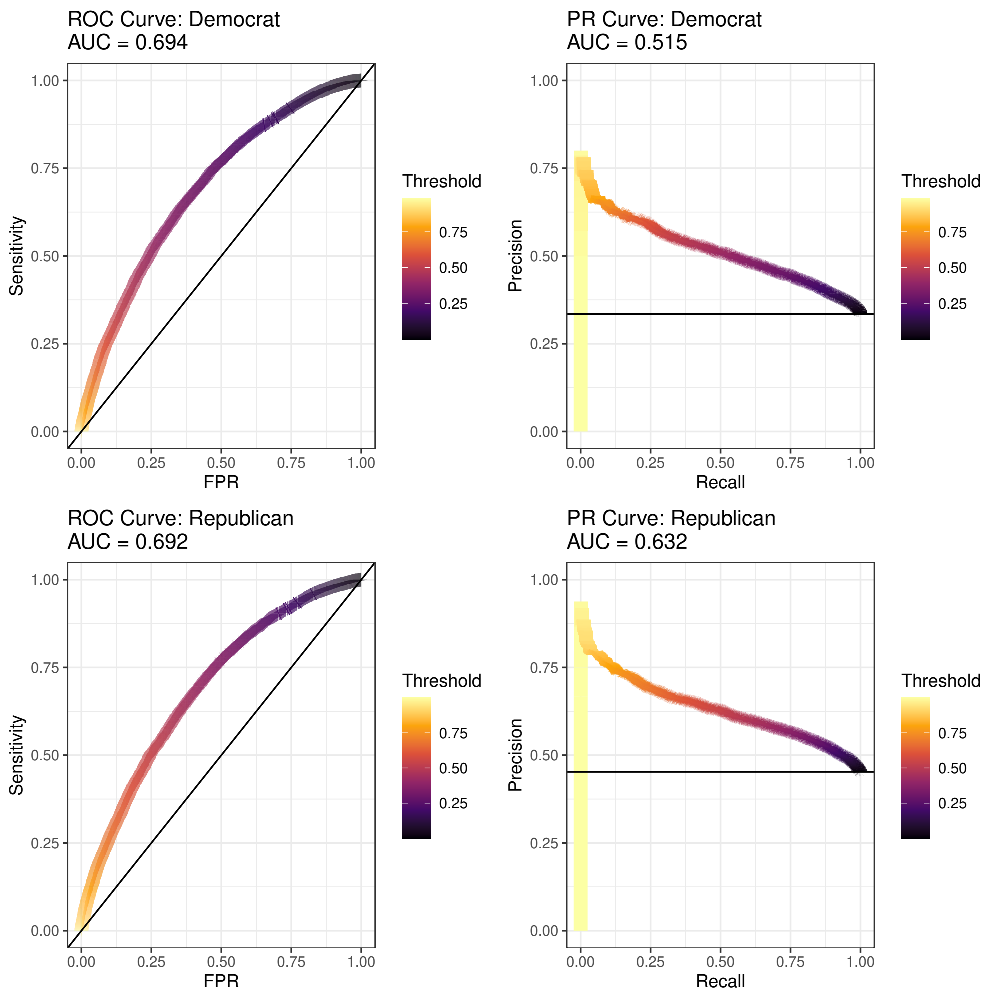
Performance in the "test'' set In figure B5, I show that the ROC and PR curves for the classifier in the "test'' set, while imperfect, substantially outperforms the naive classifier for both Republicans and Democrats. And in figure B6, I show the predicted probability of being a Republican and Democrat versus the actual probability. Again, while the classifier tends to over-predict at the high and low end, there is clearly a strong relationship between predicted and actual probability of Partisanship.

Performance in the "validation'' set In figure B7, I show the ROC and PR curves for this classifier in the "validation'' data: the 1874 biographical records linked to Union Soldiers. These were never used in training the classifier. Despite that, and despite there being substantial differences in the fraction who are Republican vs Democrat (soldiers are more Republican and non-soldiers), the classifier still outperforms the naive "no-skill'' classifier. And again, the predicted probability of partisanship is strongly related with actual probability of partisanship (Figure B8). Like the test set, though, the relationship isn't one to one. It should be noted that that the classifier does a worse job at predicting Democrats among veterans: veterans predicted with to be Democrats with nearly 100 percent probability are only actually Democrats less than 70 percent of the time. This is likely reflecting the effects I find that people were converted to being Republican.

Performance in the "aggregate census'' set This provides evidence that the classifier does a fairly good job predicting partisanship in 1874 for people whose demographic data was measured in 1874. But does it do a good job predicting partisanship during or before the Civil War, when selection into service and, potentially, combat experiences took place?

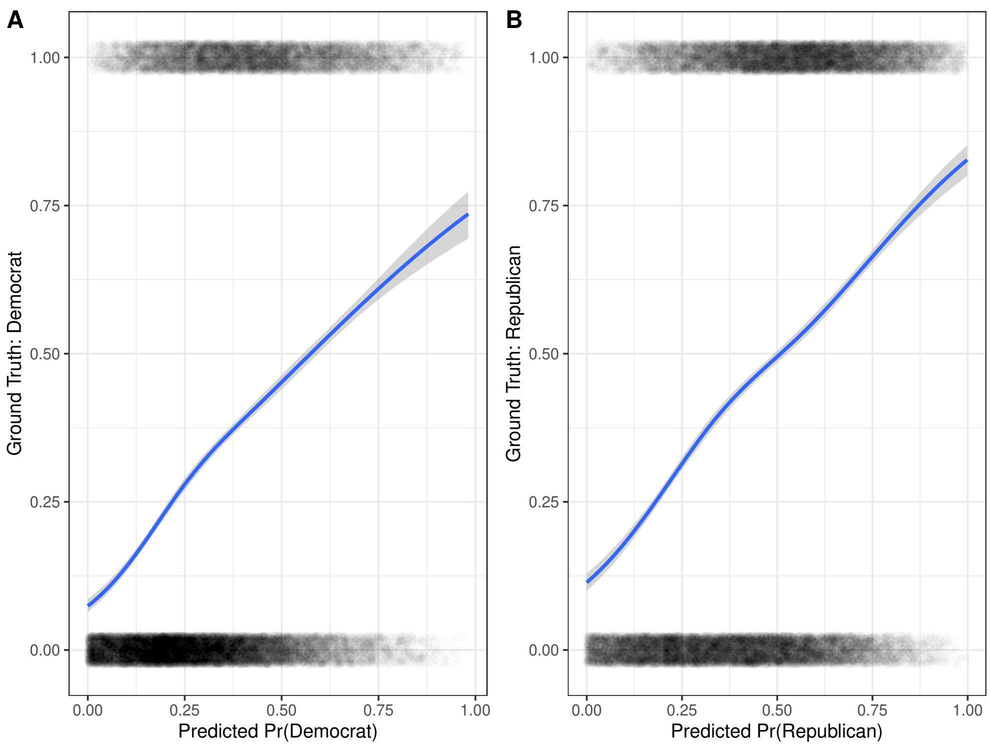
To assess this, I aggregate the predicted probabilities of Democratic and Republican partisanship for men eligible to vote in the 1860 elections and living the 9 Indiana counties,

Figure B5: AUC/PR Curves for “Test” Set



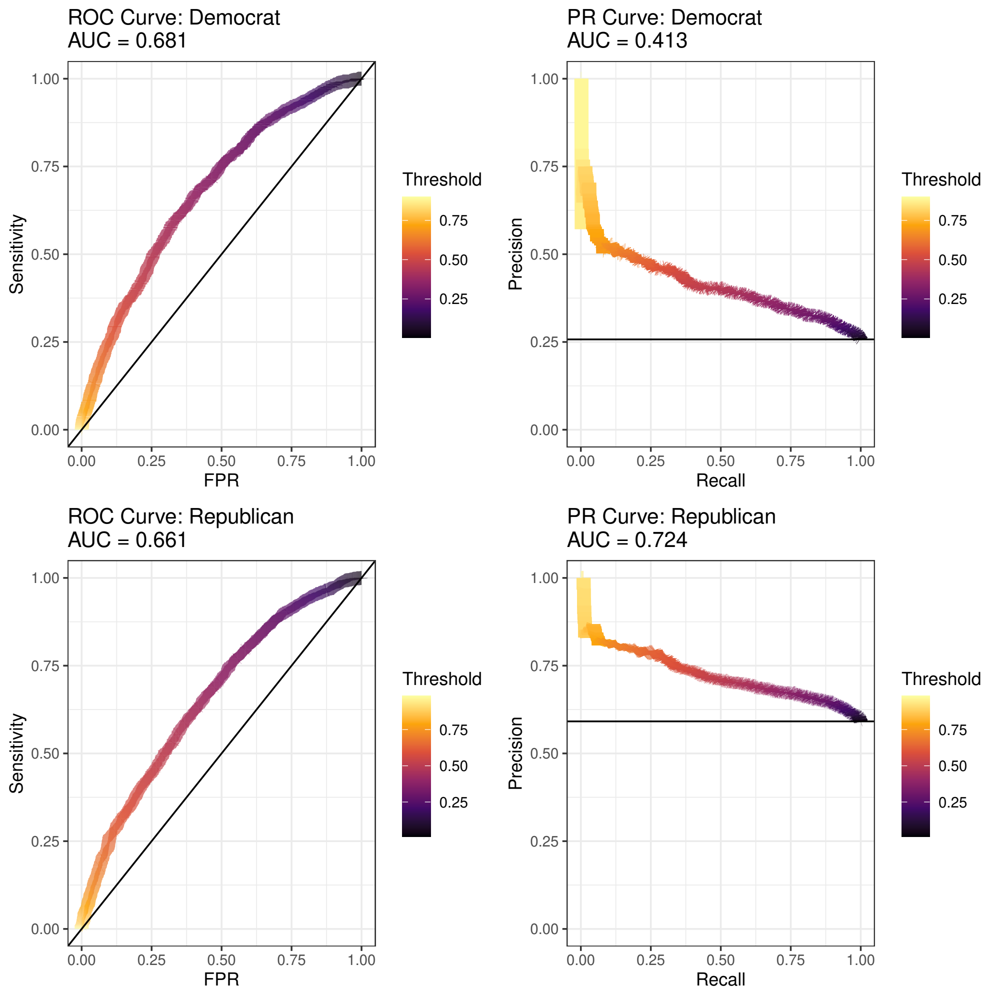
This figure shows the ROC and PR curves for the fastText classifier of Republican and Democratic partisanship in 1874, for the “test” set or cases held out in each fold when training the classifier.

Figure B6: Actual Probability of Partisanship vs. Predicted Probability of Partisanship: “Test” Set



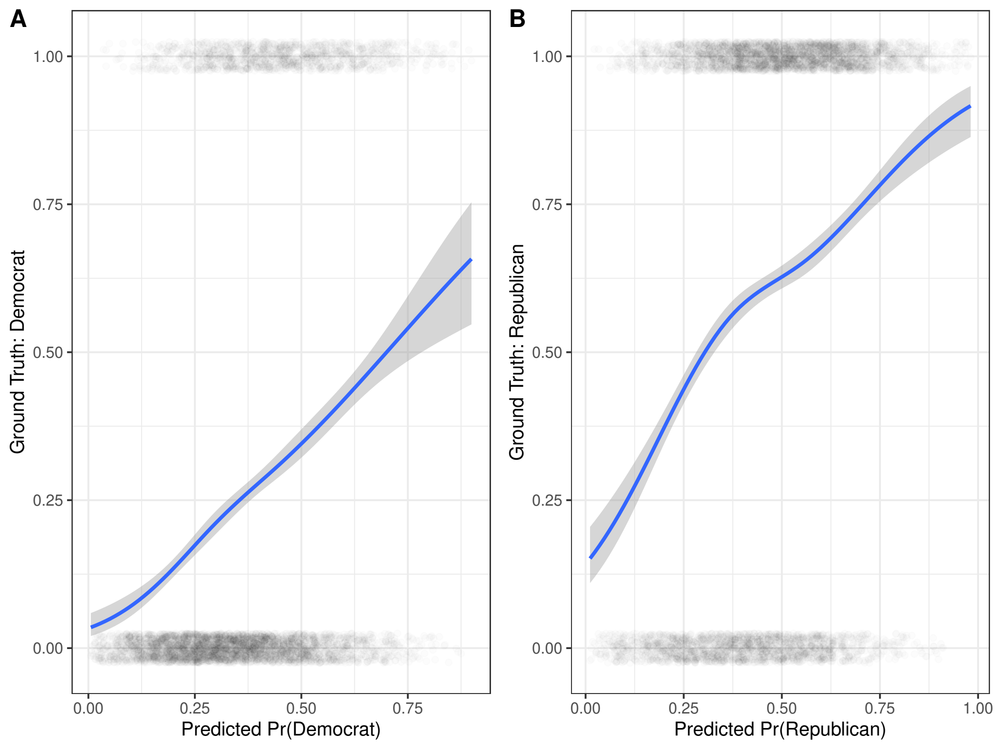
This figure shows the calibration curve of predicted partisanship vs. actual partisanship from the fastText classifier of Republican and Democratic partisanship in 1874, for the “test” set or cases held out in each fold when training the classifier.

Figure B7: AUC/PR Curves for "Validation" Set



This figure shows the ROC and PR curves for the fastText classifier of Republican and Democratic partisanship in 1874, for the "validation" set or cases linked to veterans and not used in training the classifier.

Figure B8: Actual Probability of Partisanship vs. Predicted Probability of Partisanship: "Validation'' Set

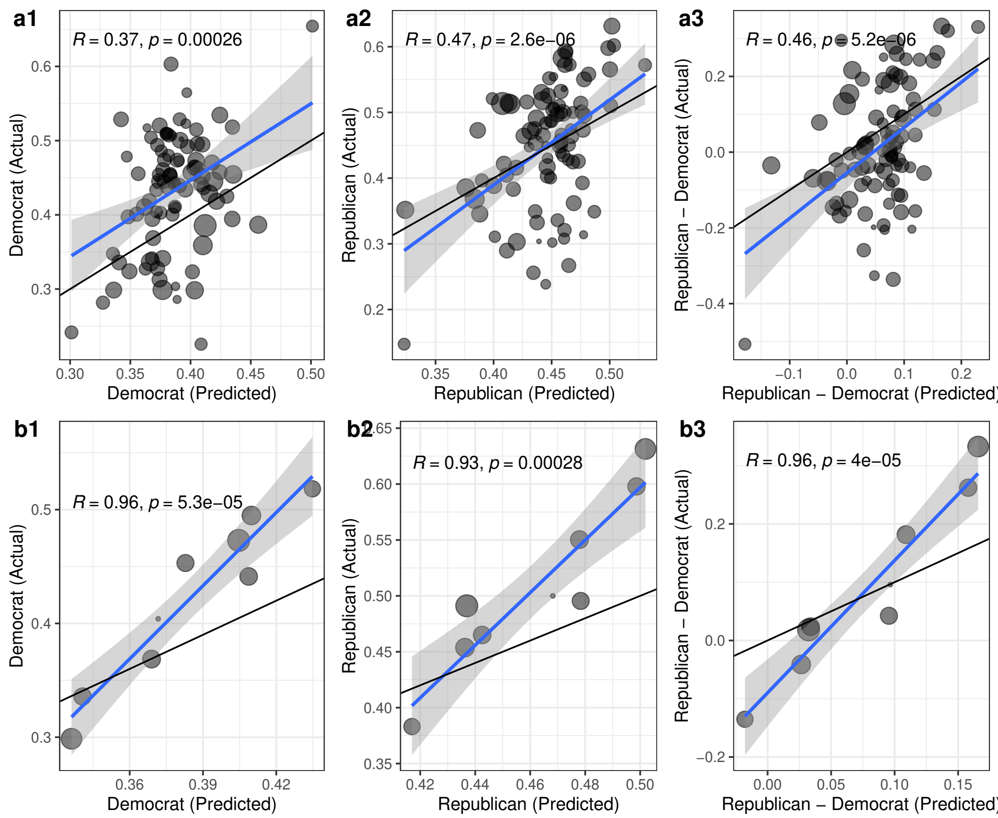


This figure shows the calibration curve of predicted partisanship vs. actual partisanship from the fastText classifier of Republican and Democratic partisanship in 1874, for the "validation'' set or cases linked to veterans and not used in training the classifier.

according the 1860 Census. I then take the average Republican probability, average Democratic probability, and the difference of these two for all men in each county (and townships in 4 counties).

In figure B9, I show the correlation between the predicted partisanship of these 9 counties in 1860 and the actual 1860 vote share for Democrats and Republicans. While, again, the predicted probability does not predict voteshare at a one-to-one level, the correlation is nearly 1. This is remarkable given that the classifier was not provided any county labels.[3]

Figure B9: Predicted vs. Actual 1860 Voteshare for 9 Indiana Counties



This figure shows the calibration curve of predicted partisanship vs. actual partisanship from the fastText classifier of Republican and Democratic partisanship in 1874, for the Indiana Counties in the 1860 Presidential election. Predicted partisanship was obtained by applying the fastText classifier to all voting-eligible males in each county and taking the mean probability of Democratic and Republican partisanship.

I repeat the same procedure using township level returns in Hendricks, Henry, Montgomery, and Morgan counties. These results are reported in Figure B10. Again, while the relationship is not one-to-one, predicted partisanship of townships in 1860 is correlated with the difference in Democratic and Republican vote share at nearly 0.8. Clearly, the model generates predictions of partisanship that correspond to real pre-war differences.

[3]The correlation between predicted partisanship and voteshare is weaker for all Indiana counties, but still highly significant.