Democratic Peace, Piece by Piece

Ransi Clark
Caltech

Jonathan N. Katz
Caltech

October 2025

Abstract

Democratic peace is an enduring empirical observation in international relations. However, the causal mechanism is disputed. Some scholars say that democracy's causal effect is confounded by other phenomena such as economic liberalization. Others propose that unobservable factors such as shared values among dyads are what drives the peacefulness among democracies. These unobservable factors also introduce pernicious biases. One such unobservable is the hidden belligerence among dyads. Some dyads are more conflict-prone than others, and most dyads are, even among the contiguous dyads, never engage in conflict. The mixing of these two types of dyads lead to both attenuation biases and selection biases. Attenuation biases arise when dyads that have never fought overwhelm the average treatment effect and weaken the signal. Selection biases arise when countries selectively democratize based on whether or not they have a bloodfeud with their neighbors. The solution we propose is to proxy for the baseline risk of conflict of a dyad using an indicator on whether the member countries had fought in the lead up to democratization, and then only compare dyads that have the same baseline risk of fighting, and report treatment effects conditionally on the risk strata. Disaggregating in this way insulates against selection on baseline and attenuation, mimicking a differences-in-differences type analysis. For countries that had a baseline risk of fighting, we recover a democratic peace effect two times the magnitude for countries that had no baseline risk. To recover these estimates, we use a dynamic algorithm that ensures that the treatment (joint democracies) and control groups (not joint democracies) are contemporaneous, have similar baseline risk, and if there are differences in observable characteristics these are also weighted for. Software that implements the algorithm is available.

1 Introduction

Democratic peace, the observation that democracies fight each other less than other types of joint regimes, is an established fact in empirical international relations. The joint occurrence of these two political phenomena, however, does not necessarily mean democratization causes inter-state peace. One threat to causal identification is that democratization requires certain pre-conditions that also enable peace. Economic growth caused by more frequent international trade flows is one such mechanism. Economic growth emboldens the middle class who finds wars costly because of the gains to be had from trade. A growing middle class also frequently demands more enfranchisement. Under such a mechanism, the impetus for peace was growth, rather than democracy. War battered states may lack resources to organize elections. In such cases, peace is a pre-condition for democracy.

If there are covariates that can account for such confounding, a causal quantity can be obtained by specifying a functional form that accounts for the covariates, or restricting comparisons to country pairs (dyads) that are similar in covariates (selection on observables). In the militarized conflict data we use the observation of dyadic democratic peace is generally robust to the inclusion of covariates such as per capita GDP, and derived quantities such as Composite Index of National Capability. But this is not uniformly supported. Gartzke (2007) finds that once a measure of financial openness is included, the democratic peace effect disappears. Imai and Lo (2021) takes a much different approach. Analogizing the democratic peace problem to the association between smoking and lung cancer, they ask how prevalent a confounder must be to explain the association between joint democracy and dyadic peace. They find that a confounder must be 47 times more prevalent in joint democracy dyads than others, and no confounder currently identified in the literature meets this threshold.

An even more difficult challenge to causality comes from unobserved factors. One manifestation of this is in the claim that democracy and inter-state peace are both ultimately due to shared culture. Democracies are reluctant to fight each other because they see them as culturally similar. Since culture cannot be captured quantitatively none of the above solutions are helpful. One methodological fix to this problem is to account for the past history of the dyad. Beck, Katz, and Tucker (1998) suggests fixed effects on spell duration (time since last conflict) to account for the past history of the dyad and any unobservable factors. Green, Kim, and Yoon (2001) suggests the inclusion of a dyadic fixed effect to control for unobservables, in addition to fixed effects on duration. Application of dyad-level fixed effects within a logit regression is, however, likely to cause large sample losses because a fixed effect logit regression can only estimate effects from the sample of countries that show variation

in the outcome. In international conflict data most dyads do not fight each other, which leads to a large portion of the sample to be lost, thus the estimates only come from the sample that transitioned to democracy within the analyst's observation window (Beck and Katz 2001).[1]

Such large sample losses do not arise under a linear fixed effects model. However, with a binary treatment, fixed effects are still a concern because dyads that never democratize are difference out and do not contribute to the treatment effect. Selection effects like this can attenuate the treatment effect if dyads that have a country that never democratizes are more belligerent than dyads with countries that democratize at some point. [2]

One may also consider that there are unobservable time related factors that increase or decrease the general peaceability regardless of regime types, such as World Wars, and be tempted to control for these unobserved temporal factors, using another fixed effect on year, in addition to dyad fixed effects. But recent econometric advances show that in a staggered treatment setting such as this, where democratization occurs at different years, dual fixed effects regressions use already democratized countries as control units for later democratizing countries. If earlier democratizing dyads are more peaceful than the dyads that never democratize then the estimate derived for these later democratizing countries will be biased downwards.

A major cause of attenuation is the over-inclusion of dyads that are never likely to fight, either because they do not have the opportunity, willingness, or both. Even once filtered to politically relevant dyads[3], many of the dyads never fight. This has the effect of diluting the signal of democratic peace. One way to mitigate this problem is to limit the sample to dyads that fight at all, but doing so also restricts the causal interpretation we can make.[4]

In summary, the methodological challenge then is to account for both observable and unobservable dyadic and time factors when the outcome is a zero inflated and binary. With a continuous outcome, researchers can use a variety of methods that conform to both selection on observables, and selection on unobservables assumptions. These include differences-in-

[1] In Green, Kim, and Yoon (2001)'s sample, out of the 3075 dyads observed only 198 have longitudinal variation, meaning the rest of the 2877 were irrelevant in the logit fixed effects model. Dyads who are always at war will also be removed for lack of variation, even though this is never practically observed.

[2] This type selection is mitigated with a continuous indicator of joint democracy. But such a model assumes that democratic peace effects increase monotonically on the least democratic member's democracy score.

[3] Dyads where members are contiguous by land or dyads where one or both members are major powers as defined by the COW dataset.

[4] Some scholars suggest modelling this problem as one of zero-inflation, where dyads are drawn from either a population that does not have opportunity to fight and therefore does not fight, and another population where there is opportunity to fight and may or may not fight. Recovering such models with binary outcomes are usually difficult, particularly without access to a rich vector of observable factors that can distinguish between opportunity and the willingness to fight.

differences, synthetic controls, and other localized causal tools. If we had access to a continuous variable that encodes the baseline propensity to fight, then the pre-transition value of this can be differenced from the post-transition value of the outcome to debias any imbalance in baseline risk between joint democracies and other regime types. When the outcome is binary, and zero most often, many of these methods fail to capture baseline information adequately.

In recommending the solution, we must first observe that the focus on a causal estimate that is averaged across all baseline risks obscures useful causal content. When the peace effect of dyads that have high risk of fighting pre-democratization are average with dyads that have little risk of fighting pre-democratization, the second effect of much smaller magnitude will overwhelm the first. Yet, if causal effects are disaggregated by baseline risk, we can interpret the democratic peace effect specific to this baseline risk. To use an analogy from medicine, a clinician that wants to estimate the preventative benefit of a drug on a rare disease, usually prefers to restrict his subjects to those who are prone to the rare disease. A clinical trial done on the entire population will underestimate the preventative benefit of such a drug, because the baseline incidence of the disease among the entire population is small.

Baseline stratification has several other advantages. For one, it also accounts for selection effects of the type where countries selectively democratize based on their past history of conflict. For another, it helps demonstrate the robustness of the democratic peace theory: if the theory holds not just in aggregate, but for all baseline risk strata, then we can be more confident about the causal role of democracy in peace.

Our recommendation is to restrict comparisons to dyads that have similar baseline risk. Baseline risk is measured by whether the dyad's countries fought each other within a window of time in the lead up to the transition year, say, past twenty years. This criteria gives us two risk strata, one which we call high risk, the other low risk. To obtain a joint democracy dyad's peace effect, supposing that the dyad becomes a joint democracy in 1960, we look back 20 years (can be varied) to determine its risk category. If the dyad has had a conflict within 1940-1960, when deriving its peace effect, we compare to non-democratic dyads that have also fought at least once within 1940-1960. In addition to keeping comparisons within risk levels, this ensures comparisons are kept contemporaneous, the effect we expect to achieve with a year fixed effect.

Our measure of baseline risk is approximate. Some countries fight more regularly than others even when restricted a look-back window. To ensure robustness we vary this window. Varying represents a tradeoff, if the window is taken to be too long, it may misclassify several low risk dyads to be high risk attenuating the democratic peace effect among the high risk. If the window is taken to be too short, our sample would be too small. We do a series

of simulations to consider the effect of varying window lengths. Generally, the higher the number of low risk dyads in the sample, the more attenuation is to be expected. Shortening the window slows the attenuation as more and more low risk dyads enter the analysis. From a split population point of view, the better the separation between the zero-inflation portion and non-zero portion, the attenuation is mitigated.

This sequestration to risk groups is difficult to implement in a regression framework, since we want to sequester comparison sets on pre-transition baseline information only.[5] Therefore, we estimate these effects non-parametrically in a two-step framework. The first step calculates each joint democracy dyad's peace effect at all of its post-democratization periods. As discussed above, these estimates keep comparisons (the joint democracy dyad vs all other non-democracy dyads) contemporary and within the same risk category. These component estimates can then be aggregated into their risk strata, and even further to give an overall differences-in-differences like effect of democratic peace. The estimator allows further decomposability to investigate what dyad contribute most to the effect. Since the method requires that treatment and control groups be exactly identified, we use a binary democratization indicator.[6]

Doing so we find that the democratic peace effect for high baseline risk dyads is more than twice that of low baseline risk dyads. Further disaggregation by time of democratization shows that the peace effect was strongest for dyads that transitioned to joint democracy in the 1940s. This implies that a major driver of democratic peace are European countries that transitioned or returned to democracy after World War II.

In the next section, we describe the type of data generating process that we imagine generates conflict data. We analyze treatment effects recovered by regressions and our proposed method to demonstrate how attenuation and selection biases arise. After this, we introduce our estimation algorithm. Then, we implement this algorithm on dyadic conflict data with a binary joint democracy indicator.

[5]When a dynamically updating indicator of fighting within a window is included in the regression specification, this also affects the post-transition period.

[6]Many models use continuous indicators of joint democracy. But the interpretation of these models is the marginal effect of democratic peace, rather than the treatment effect. Since we are specifically trying to recover a treatment effect, we need a binary indicator.

2 Hidden belligerence

We now introduce the data generating process, from which we draw simulations to demonstrate the behavior of common estimators. We also use this to define the causal targets of interest.

We first define a variable B_i for each dyad i which is the belligerence of the dyad. To keep our calculations tractable, we say, B_i is binary, where $B_i = 1$ is a belligerent dyad and $B_i = 0$ is not. The binary indicator $D_{i,t}$ is the joint democracy indicator. If $D_{i,t} = 1$ the dyad i is a joint democracy in year t. The binary outcome $Y_{i,t}$ encodes the occurrence of conflict and is a function of both B_i and $D_{i,t}$. For illustrative purposes, we will drop the subscripts on B_i , $D_{i,t}$, and $Y_{i,t}$ referring to them simply as B, D and Y, respectively.

Let the number of belligerent dyads in the sample be N_1 and the number of non-belligerent dyads to be N_0. Then the parameter p_B can be calculated as the ratio $\frac{N_1}{N_0+N_1}$. An fall in p_B represents that the sample has more and more non-belligerent dyads. Similarly, let p_D be the ratio of joint democratic dyads to all other dyads in the sample.

Also notate the dependence structure between D and B with $\delta = a_1 - a_0$, where $a_1 = \mathbb{P}(D = 1|B = 1)$ and $a_0 = \mathbb{P}(D = 1|B = 0)$. If D and B are independent, or the probability that a belligerent dyad will become a democracy is the same as the probability that a belligerent dyad will become a democracy, $\delta = 0$.

Though B is not observed, our interest in the causal effect of democratization is conditional on B. Under a binary B, there are three causal targets: the effect of democratization for belligerent dyads $ATT(B = 1)$, the effect of democratization for non-belligerent dyads $ATT(B = 0)$, and the combined causal effect of democratization ATT.

Assume the following data generating process for outcome Y,

$$m(D, B) = \Pr(Y = 1 \mid D, B) = \Lambda(\alpha_0 + \alpha_D D + \alpha_B B), \quad \Lambda(t) = \frac{e^t}{1 + e^t}.$$

We need that more belligerent dyads are more likely to fight regardless of regime type, and therefore assume that $\alpha_B > 0$. We also assume that non-belligerent and non-democratic states engage in conflict rarely enough so that $\alpha_0 < 0$.

Since both D and B are binary, we can define our causal targets $ATT(B = 1)$ and $ATT(B = 0)$ in terms of $m(D, B)$.

$$ATT(B = 1) = m(1, 1) - m(0, 1)$$
$$ATT(B = 0) = m(1, 0) - m(0, 0)$$

Although, not of causal interest, we also define the effect of B on D for easier demonstration.

$$\Delta(D=1) = m(1,1) - m(1,0)$$
$$\Delta(D=0) = m(0,1) - m(0,0)$$

Since $\alpha_B > 0$, both $\Delta(D=1)$ and $\Delta(D=0)$ are positive.

Observation 1: If $\delta = 0$, $\widehat{ATT} = p_B * ATT(B=1) + (1-p_B) * ATT(B=0)$.

If D and B have zero covariance (or equivalently $\delta = 0$) the estimate $\widehat{ATT}$ recovered from a regression of Y on D, is a convex combination of $ATT(B=1)$ and $ATT(B=0)$ with the parameter p_B, the ratio of belligerent dyads in the sample.

This can be expressed as,

$$\widehat{ATT} = \mathbb{E}[m(1,B)|D=1] - \mathbb{E}[m(0,B)|D=0].$$

To expand these terms, we appeal to the law of iterated expectations.

$$\begin{aligned}
\mathbb{E}[m(1,B)|D=1] &= \sum_{b\in\{0,1\}} m(1,b)Pr(B=b|D=1) \\
&= m(1,1) * \mathbb{P}(B=1|D=1) + m(1,0) * \mathbb{P}(B=0|D=1) \\
&= m(1,0) + \mathbb{P}(B=1|D=1)(m(1,1) - m(1,0)) \\
\mathbb{E}[m(0,B)|D=0] &= \sum_{b\in\{0,1\}} m(0,b)Pr(B=b|D=0) \\
&= m(0,1) * \mathbb{P}(B=1|D=0) + m(0,0) * \mathbb{P}(B=0|D=0) \\
&= m(0,0) + \mathbb{P}(B=1|D=0)(m(0,1) - m(0,0))
\end{aligned}$$

Under independence of B and D, $\mathbb{P}(B=1|D=1) = \mathbb{P}(B=1|D=0) = p_B$, and $\mathbb{P}(B=0|D=1) = \mathbb{P}(B=0|D=0) = 1-p_B$.

Substituting the definitions of $ATT(B=1)$ and $ATT(B=0)$, we can simplify, $\widehat{ATT}$ to be a convex combination of these terms.

$$\begin{aligned}\widehat{ATT} &= \mathbb{E}[m(1,B)|D=1] - \mathbb{E}[m(0,B)|D=0] \\ &= p_B * (m(1,1) - m(0,1)) + (1-p_B) * (m(1,0) - m(0,0)) \\ &= p_B * ATT(B=1) + (1-p_B) * ATT(B=0)\end{aligned}$$

Even though we assume that the democratization parameter α_D does not depend on B, the causal effects $ATT(B=1)$ and $ATT(B=0)$ can differ drastically. The smaller that α_0 is or the larger that α_B is the difference between the two effects grow larger.

Because p_B is usually small, overall effects are closer to $ATT(B=0)$ than $ATT(B=1)$, which means that the overall effect obtained from a regression of say Y on D is likely to be very close to zero. In other words, most of the information in the average treatment effect comes from non-belligerent dyads. However, this effect is less interesting than the effect democratization has on belligerent dyads. To return to the example of the preventative treatment of a rare disease, our interest should be on the prevention's effect on those with a propensity to get the disease. A population level treatment effect would obscure the preventative benefit of the treatment.

If the dependence of D and B, which we denote using δ, is different from zero, an estimate for ATT obtained by a regression of Y on D is biased. Denote this estimate as $\widehat{ATT}$. This type of dependence can either be caused by belligerent dyads choosing into democratization ($\delta > 0$) or a reluctance of belligerent dyads to democratize ($\delta < 0$).

Observation 2: If $\delta > 0$, $\widehat{ATT}$ is upward biased compared to the ATT obtained under independence. If $\delta < 0$, $\widehat{ATT}$ is downward biased compared to the ATT obtained under independence.

First we recast $\mathbb{P}(B=1|D=1)$ and $\mathbb{P}(B=1|D=0)$ in terms of p_B, p_A, a_1, a_0. From the law of iterated expectations,

$$p_D = a_1 p_B + a_0(1-p_B)$$

Using Bayes' rule

$$\begin{aligned}\mathbb{P}(B=1|D=1) &= \frac{a_1 p_B}{p_D} \\ \mathbb{P}(B=1|D=0) &= \frac{(1-a_1)p_B}{1-p_D}\end{aligned}$$

Using the definition of p_D, their difference can be expressed in terms of δ

$$\mathbb{P}(B=1|D=1) - \mathbb{P}(B=1|D=0) = \frac{p_B(1-p_D)}{p_D(1-p_D)} = \frac{p_B(1-p_B)\delta}{p_D(1-p_D)}$$

Taking differences of $\mathbb{E}[m(1,B)|D=1] - \mathbb{E}[m(0,B)|D=0]$ and substituting the appropriate definitions of $\mathbb{P}(B=1|D=0)$ gives,

$$\begin{aligned}\widehat{ATT} &= ATT(B=0) + \frac{a_1 p_B}{p_D} * (ATT(B=1) - ATT(B=0)) + \frac{p_B(1-p_B)\delta}{p_D(1-p_D)}\Delta(D=0) \\ &= \frac{a_1 p_B}{p_D} * ATT(B=1) + \left(1 - \frac{a_1 p_B}{p_D}\right) * ATT(B=0) + \frac{p_B(1-p_B)\delta}{p_D(1-p_D)}\Delta(D=0)\end{aligned}$$

Now the $\widehat{ATT}$ recovered from the regression of Y on D is a convex combination of $ATT(B=1)$ and $ATT(B=0)$ as well as an additional term that is always positive if $\alpha_B > 0$. For fixed a_1 and p_D as p_B falls the weight on $ATT(B=1)$ falls. As p_B increases towards 1 or decreases towards 0, the weight on the third term falls too.

The dependence between B and D affects the recovered estimate through this third term. Notice that if $\delta = 0$ the third term drops off. Also $\frac{a_1 p_B}{p_D}$ simplifies to p_B when $\delta = 0$ since $a_1 = a_0$. This gives us the same expression of $\widehat{ATT}$ as we obtained previously.

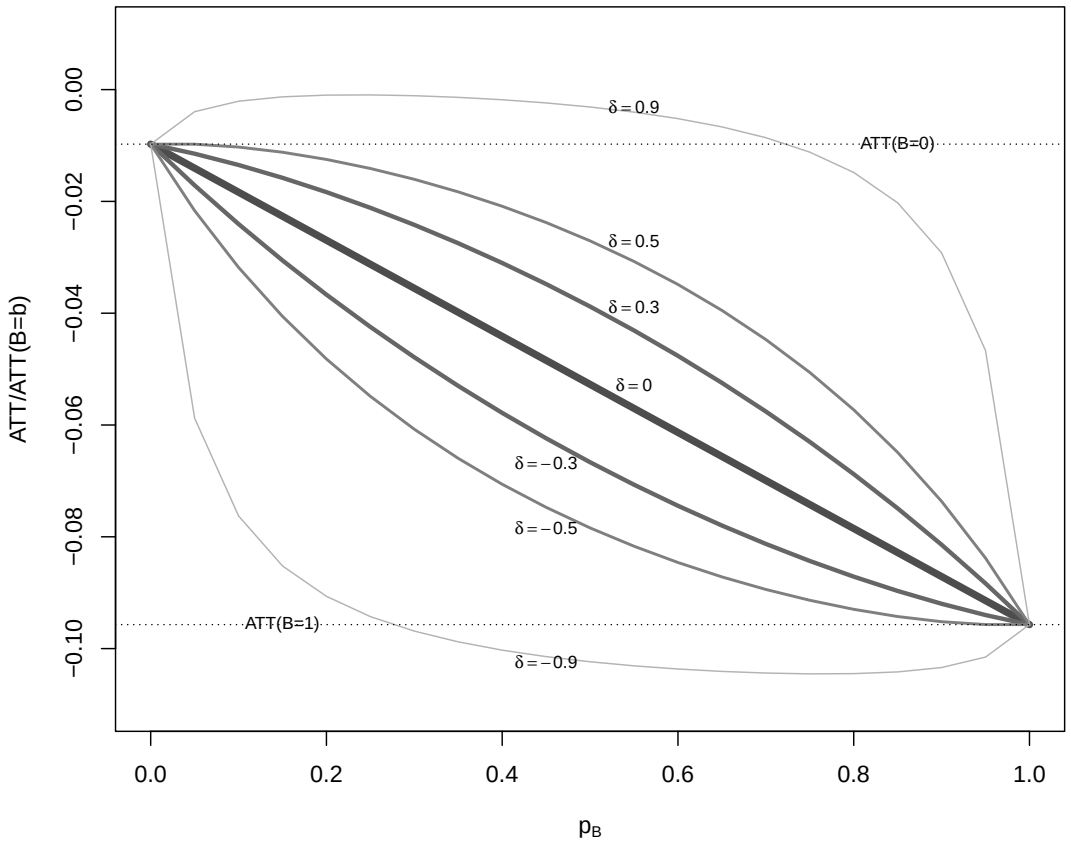
If $\delta > 0$, the additive term is positive and upward biases the recovered estimate. If $\delta < 0$, the additive term is positive and downward biases the recovered estimate.

Figure 1 illustrates the $\widehat{ATT}$s recovered by a regression for fixed α_0, α_B, α_D while varying p_B and δ.

The parameter α_0 was fixed at -4.5, implying that for a non-belligerent non-democratic dyad the probability of conflict is 0.01. The parameter α_B was fixed at 2.4, implying that a non-democratic belligerent dyad had a probability of conflict of about 0.09. The α_D parameter was fixed at -2, implying that a non-belligerent democratic dyad had a probability of conflict of about 0.002. These parameters were chosen to reasonably represent the observed probability of conflict in the data, even though we do not know what a belligerent dyad is and what is not.

Even though the democracy parameter α_D in the model is the same for a belligerent or a non-belligerent dyad, the $ATT(B=0)$ implied by the model (-0.009) is 10 times as smaller than $ATT(B=1)$ implied by the model (-0.09). When independence of B and D is assumed, as in Observation 1, the $\widehat{ATT}$ is a convex combination of $ATT(B=0)$ and $ATT(B=1)$ with parameter p_B. The more non-belligerent dyads there are in the sample, the closer p_B is to zero and smaller is $\widehat{ATT}$. In just the sample of 2420 politically relevant dyads, 1681 of

Figure 1: Recovered treatment effect $\widehat{ATT}$ for varying δ and p_B



Note: The parameter p_B is the proportion of belligerent dyads in the sample. The parameter δ is the difference in the proportion of belligerent dyads that are joint democracies and proportion of non-belligerent dyads that are joint democracies. The sign of δ is the same direction as the covariance between B (belligerence) and D (democracy). The plot is generated using the parameters $\alpha_0 = -4.5$, $\alpha_B = 2.4$ and $\alpha_D = -2$. The plot shows that as as p_B falls, the regression estimate $\widehat{ATT}$ tends to $ATT(B = 0)$. As δ falls towards -1, the regression estimate $\widehat{ATT}$ is downward biased, and in the converse direction is upward biased.

the dyads were never in a dispute. This puts the upper bound for p_B at 0.3. If all dyads, regardless of their political relevance, are included the estimate falls further towards zero.

Even at this small p_B, the dependence structure implied by $\delta = a_1 - a_0$ can further attenuate the effect if δ is positive. For a fixed p_B, the larger positive δ increases the attenuation bias. If $\delta < 0$, the effect is downward biased. Since we do not know which dyads are belligerent we cannot empirically determine δ. However, we can arrive at a rough estimate from other empirical quantities. Out of the dyads that never fought about 30% are democracies, while only about quarter of all dyads are joint democracies. So we can estimate

$\delta \approx -0.25$.

Despite the negative bias from the correlation, the estimate $\widehat{ATT}$ is still further away from $ATT(B = 1)$ than it is from $ATT(B = 0)$ due to the pre-ponderance of non-belligerent dyads in the data. If our interest is in the effect of democratization on the type of dyads that are likely to engage in a conflict $\widehat{ATT}$ recovered from a regression is an underestimate of this effect.

2.1 Proxying hidden belligerence

While our causal target may be $ATT(B = 1)$, or even $ATT(B = 0)$, we do not observe B directly. If we could we could partition the data to the two strata and estimate regressions within the strata. One approach is to consider only those dyads that have seen a conflict at least once in the sample's time window at a belligerent dyad and all others as non-belligerent dyads. Under this approach the estimate for $ATT(B = 0)$ will always be zero. This implies that this type of classification is too restrictive since even non-belligerent dyads fight, but less frequently.

A less restrictive but practicable solution is to proxy the risk strata that a dyad belongs to by classifying it into some stratum based on its conflict history. To ensure sufficient sample sizes within each stratum, we limit to just two strata. The dyad's risk stratum updates dynamically, so denote the dyad's risk stratum at time t as $S_t(h)$. A dyad belongs to the belligerent stratum $S_t(h) = 1$ if within a window of h years from time t, the dyad was observed as engaged in a conflict, and belongs to the non-belligerent stratum $S_t(h) = 0$ if the dyad was not observed in a conflict in that time window. We suppress the dependence of S on h and t from here onwards and refer to it simply as S.

Observation 3: The treatment effect $\widehat{ATT}(S = 1)$ where $S = 1$ if a conflict is observed within h periods of t is tends towards $\widehat{ATT}$ as h grows, when assuming independence of S and D that is, $\mathbb{P}(B = a|S = a, D = 1) = \mathbb{P}(B = a|S = a, D = 0)$, for all $a \in \{0, 1\}$.

In other words, assuming that the (mis)classification of B into S does not depend on democratization, the longer the look back window h, both $S-$specific ATTs tend to the $\widehat{ATT}$ recovered from a linear regression of Y on D. For a $\alpha_D < 0$ and for any finite h, this implies that the underestimate of $\widehat{ATT}(S = 1)$ worsens with h, but that the overestimate of $\widehat{ATT}(S = 0)$ improves with h.

Denote the following probabilities as π_1 and π_0.

$$\pi_1 = \mathbb{P}(S = 1|B = 1)$$
$$\pi_0 = \mathbb{P}(S = 1|B = 0)$$

These quantities are the sensitivity of S to B and the misclassification of B respectively.

Since our classification rule to determine S is a function of h, that is, a dyad seen in conflict in past h years are classified as $S = 1$, these π_1 and π_0 can be expressed in terms of h, where p_1 and p_0 are probabilities of observing conflict under $B = 1$ and $B = 0$.

$$p_1 = \mathbb{P}(Y = 1|B = 1)$$
$$p_0 = \mathbb{P}(Y = 1|B = 0)$$

Reexpressing $\pi_a(h)$ gives

$$\pi_1(h) = 1 - (1 - p_1)^h$$
$$\pi_0(h) = 1 - (1 - p_0)^h.$$

Both functions are increasing in h. That is, both the sensitivity and misclassfication rate grows in h.

Then $\widehat{ATT}(S = 1)$ can be expressed in terms of $ATT(B = 1)$, $ATT(B = 0)$, $\pi_1(h)$, $\pi_0(h)$ and p_B.

$$\begin{aligned}\widehat{ATT}(S = 1) &= \mathbb{E}[m|D = 1, S = 1] - \mathbb{E}[m|D = 0, S = 1] \\ &= w_1(1)ATT(B = 1) + (1 - w_1(1))ATT(B = 0)\end{aligned}$$

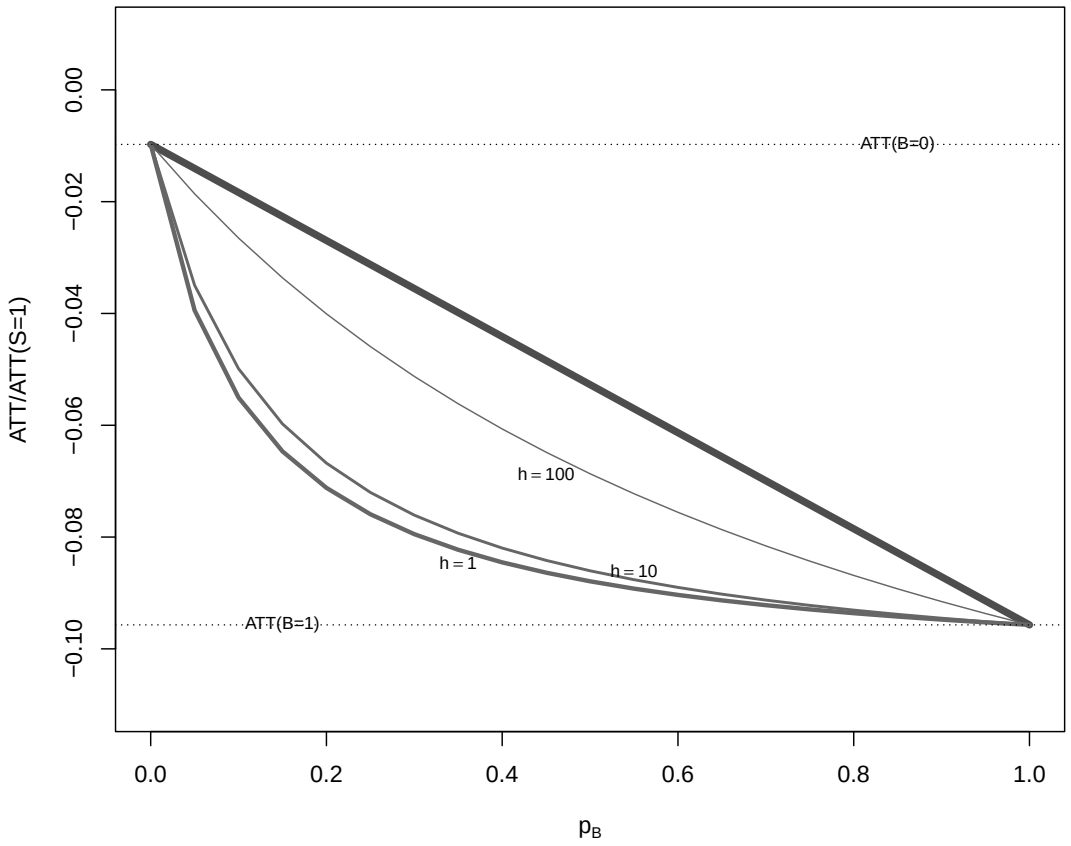
where

$$\begin{aligned}w_1(1) &= \frac{p_B\pi_1(h)}{p_b\pi_1(h) + (1 - p_B)\pi_0(h)} \\ &= \frac{1}{1 + \frac{1-p_B}{p_B}\frac{\pi_0(h)}{\pi_1(h)}}\end{aligned}$$

Since $p_1 > p_0$, the term $\frac{\pi_0(h)}{\pi_1(h)}$ grows with h, and therefore $w_1(h)$ falls with h, tending

towards p_B. When $w_1(h)$ is replaced with p_B we arrive at $\widehat{ATT}$.

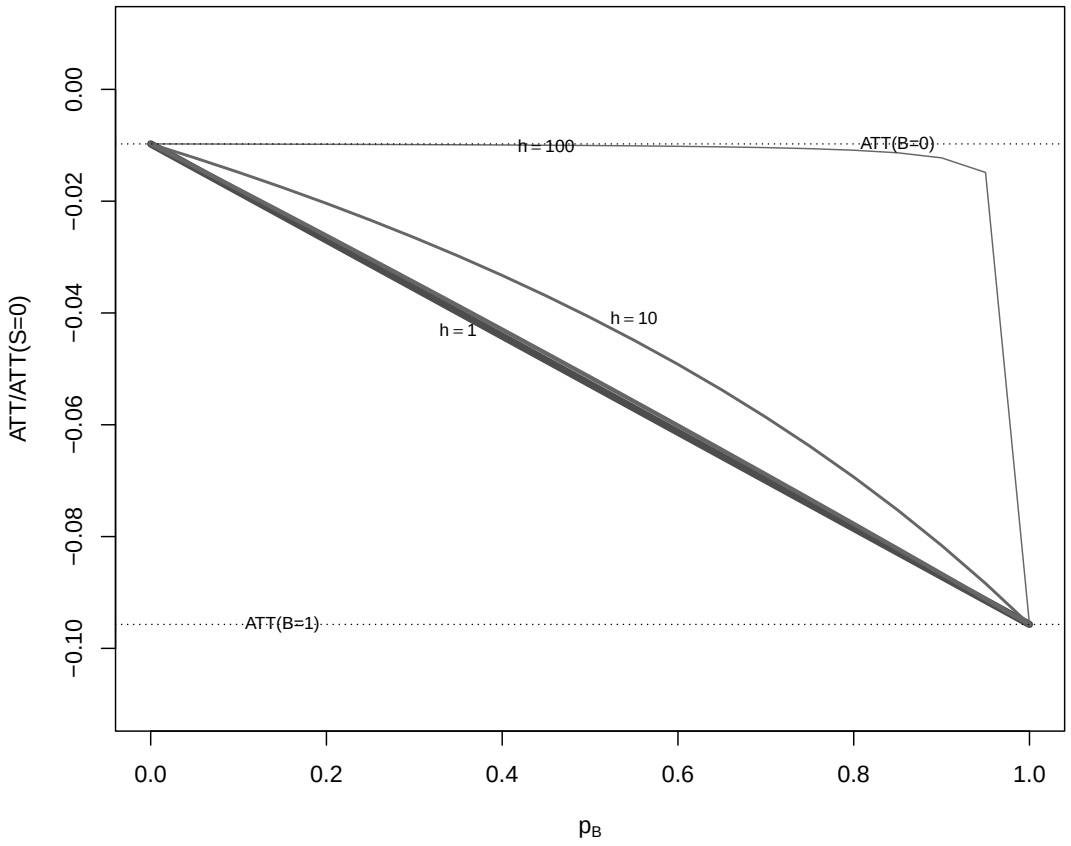
Figure 2: Recovered treatment effect $\widehat{ATT}(S = 1)$ for varying h and p_B



Note: The parameter p_B is the proportion of belligerent dyads in the sample. The parameter h is the length of the look back window. The δ is assumed to be zero. The plot is generated using the parameters $\alpha_0 = -4.5$, $\alpha_B = 2.4$ and $\alpha_D = -2$. The plot shows that as as p_B falls, the conditional estimate $\widehat{ATT}(S = 1)$ tends to $ATT(B = 0)$. As h increases, misclassification of non-belligerent dyads as belligerent dyads worsens.

Figure 2 demonstrate the estimate $\widehat{ATT}(S = 1)$ for varying look back windows h for each p_B under the data generating process used for the previous Fig. 1, assuming that δ=0 and a_1 and a_0 are interior. The estimate is closer to the true $ATT(B = 1)$ the larger p_B is. The estimates under $h = 1$ is closer to the true $ATT(B = 1)$ than the estimates under $h = 10$. As the look back window lengthen the estimates tend towards the $\widehat{ATT}$, denoted by the thick diagonal line across the figure. Under that assumed data generating process, as h increases $\widehat{ATT}(S = 1)$ is further attenuated as more and more non-belligerent dyads are misclassified as stratum 1 dyads.

Figure 3: Recovered treatment effect $\widehat{ATT}(S=0)$ for varying h and p_B



Note: The parameter p_B is the proportion of belligerent dyads in the sample. The parameter h is the length of the look back window. The δ is assumed to be zero. The plot is generated using the parameters $\alpha_0 = -4.5$, $\alpha_B = 2.4$ and $\alpha_D = -2$. The plot shows that as as p_B falls, the conditional estimate $\widehat{ATT}(S=0)$ tends to $ATT(B=0)$.

Figure 3 demonstrate the estimate $\widehat{ATT}(S=0)$ for varying look back windows h for each p_B under the same data generating process as before. An increasing the look back window brings estimates closer to the true value of $ATT(S=0)$. At the extremes of $p_B = 0$ and $p_B = 1$, the estimates coincide with the $\widehat{ATT}$.

The theoretical implication of this is that the choice of the window is a trade-off between the bias $|\widehat{ATT}(S=0) - ATT(B=0)|$ and $|\widehat{ATT}(S=1) - ATT(B=1)|$. Increasing the window length increases bias in the latter and reduces the bias in the former. Practically, however, constraint is in the number of joint democracy dyads that are classified in each strata. Even with a very large number of dyads, few dyads engage in conflict and even fewer dyads will have pre-democratic history of conflict. This makes that we must choose

a somewhat large window to get sufficient dyads for estimation purposes. We discuss this further in the results section.

3 Algorithm

Here we detail the dynamic algorithm we use to estimate $\widehat{ATT}(S = s)$. We define by G_i, the first year of joint democratization by dyad i. If the dyad is never a joint democracy, let $G_i = \max(t)$. Also denote by $S_{i,t}$, which stratum the dyad i is in at time t. As described previously, we restrict to two strata: one where a history of conflict was observed in the h window and where a history of conflict was not observed. Because for any joint-democracy dyad all the information that matters is their pre-transition stratum, we simply denote each joint-democracy dyad's stratum by $s_i = S_{i,g_i}$.

At each time t, then we can define two control sets, $C_t(S'_t = s)$, one for each strata and where $t \geq t'$. The control set $C_t(S'_t = 1)$ comprises of all the dyads that were not joint democracies at time t' and are still not joint democracies at time t. This means that over time for each strata $C_t(S'_t = s)$ is getting smaller, as dyads become joint democracies or leave the sample.[7]

The algorithm is a two-step algorithm. In the first-step, an individual treatment effect is estimated for every dyad that is observed as a joint democracy.[8] These estimates are restricted to their relevant stratum.

The second-step is an aggregation step where the first-step estimates are aggregated to produce quantities such as $\widehat{ATT}(S = 1)$ and $\widehat{ATT}(S = 0)$. These two quantities can be further aggregated to obtain an overall $\widehat{oATT}$.[9] Because we have access to all dyad-level treatment effects, we can also produce aggregates more granular than $\widehat{ATT}(S = 1)$. One quantity of interest would be $\widehat{ATT}_g(S = 1)$, which is the democratization effect of all joint democracy dyads with conflict history that jointly democratize at time g. Another quantity of interest is $\widehat{ATT}_t(S = 1)$ which is the democratization effect of all joint democracy dyads with conflict history at a particular point in time.

This first-step dyad-time level treatment effect is obtained by comparing the dyad's outcome at time t to the set of non-democracy dyads at time t that had shared the same risk stratum as the dyad i at time g_i. Its formula is as follows.

$$ATT(i, g_i, t, s_i) = Y_{i,t} - |C_t(S_{j,g_i} = s_i)|^{-1} \sum_{j \in C_t(S_{j,g_i} = s_i)} Y_{j,t}.$$

[7] Countries can leave the sample due to territorial changes.

[8] If the dyad is not observed with a history of h years, we make the assumption that the dyad did not fight for the time it was not observed. Exceptions to this rule are violent colonial detachments where even if one of the countries were not state before declaring independence this information is encoded as an observed conflict.

[9] This overall estimate is not the same as $\widehat{ATT}$ that would be obtained by a linear regression of Y on D.

where s_i is the risk stratum the dyad is classified to.

Compared to a usual differences-in-means analysis, the first step algorithm retains greater control over the composition of units. For one, the dyads in the control set must also have been observed at time g_i and cannot have entered the sample in the time between g_i and t. For another, as mentioned previously, the control set $C_t(S_{j,g_i} = s_i)$ does not accept new units over time, although it may lose units that democratize. These restrictions ensure that the comparison between the dyad i and its control set had the same baseline risk at the time of transition.

We can obtain the post-treatment dyad-level treatment effect as,

$$ATT(i, s_i) = |\max(t) - g_i - 1|^{-1} \sum_{t=g_i}^{\max(t)} ATT(i, g_i, t, s_i)$$

This quantity aggregates over the relevant joint democracy dyad's treatment effect across all its post-treatment time periods, from g_i to the maximum time of sample $\max t$.

We can then aggregate these according to the dyad's risk stratum. Suppose that for each $s \in \{0, 1\}$, S_s is the set of all joint-democracy dyad's that democratized with $s_i = s$, and $|S|_s$ is its cardinality.

$$ATT(s) = |S_s|^{-1} \sum_{j \in |S_s|} ATT(i, s_i = s)$$

The overall estimate can be obtained by aggregating regardless of the risk stratum of i.

$$oATT = (|S_0| + |S_1|)^{-1} \sum_{i \in |S_s|} ATT(i, s_i = s)$$

As mentioned previously, two other aggregates that are of interest is the democratization effect of dyads based on the time of democratization. One might theorize that early democracies are more likely to be culturally similar than those that democratize later, and therefore have a much stronger democratic peace effect. We could test the empirical content of such an theory by aggregating within time g. Supposing that $S_{s,g}$ is the set of all joint democracy dyads that democratized at time g within risk strata s.

$$ATT(s,g) = |S_{s,g}|^{-1} \sum_{i \in |S_{s,g}|} ATT(i, s_i = s)$$

Aggregates at the time level can be obtained as follows, where the set $S_{s,t}$ includes all dyads that are joint-democracies at time t and democratized with risk strata s.

$$ATT(s,t) = |S_{s,t}|^{-1} \sum_{i \in |S_{s,t}|} ATT(i, g_i, t, s_i = s)$$

3.1 Inclusion of covariates

Thus far, our discussion has abstracted away from observed covariates. Usually covariates are included as additional regressors in a regression. This practice, however, assumes a restrictive functional form for the way covariates affect Y. For one, it assumes that covariates have the same effect on Y in all years. This is a strong restriction because in periods of World Wars, major powers were more active than in other periods, implying that the major powers' effect on the probability of war differs in time. Moreover, the linear functional form also assumes that covariates have the same effect on Y regardless of the dyad's baseline risk. Given, our previous discussion, it is likely that the effect of other covariates on the outcome Y is also a function of unobservables such as belligerence, and as the number of non-belligerent dyads overwhelm the sample their effects also attenuate.

Instead of using covariates as regressors, we use them for propensity weighting when calculating the first-step quantities. In this way, the control group is weighted to best reflect the observable characteristics of each joint-democracy dyad. As an example, assume that our only other covariate is GDP per capita, and that the joint democracy dyad in question is one where the minimum of the GDP per capita in the dyad is relatively high in the sample, such as USA-Canada. Then the control group for such a dyad is weighted so that non-democratic dyads with high GDP per capita (of poorest dyad member) are weighted higher than the non-democracy dyads with low GDP per capita (of poorest dyad member). This ensures that comparisons are not only made within the same baseline risk strata and same time, but also that within those risk strata comparisons are similar in other respects.

Suppose that for each joint-democracy dyad and its relevant control group, a propensity score is obtained by a logit regression of the dyad's democracy status on a vector of covariates, $e_i(X_j)$. The predictive model $e_i(\cdot)$ that produces the propensity score is specific to each i. These propensity scores can then be used to form an inverse propensity weight w_j for each

non-democratic dyad j in the control set.

$$w_j = \frac{e_i(X_j)}{1 - e_i(X_j)}$$

This can be normalized in such a way to sum to 1 across all j, so that

$$\tilde{w}_j = \frac{w_j}{\sum_j w_j}$$

Now, instead of uniformly weighting all non-democracy dyads in the control set, they can be weighted by their normalized inverse propensity weights, as below.

$$ATT(i, g_i, t, s_i) = Y_{i,t} - \sum_{j \in C_t(S_{j,g_i} = s_i)} \tilde{w}_j Y_{j,t}.$$

3.2 Inference

While the estimation of treatment effects at a dyad level allows us more flexible aggregation across various important attributes of a treated unit, inference is challenging in this one-treated-unit setting. Usual inferential procedures such as those using covariance matrices will undercover because there are insufficient treatment units (joint democracy dyads) to approximate the variance of the treated sample. We instead use a predictive inference procedure named the Jackknife+ which considers the single democracy as having drawn from the same distribution as the control units and approximates its variance using that of the control units (Barber et al. 2021). Under the assumption that the variances across the sample is similar, undercoverage is accounted for.

4 Results

The militarized interstate disputes data was obtained from Gibler and S. V. Miller (2024). Democracy data was obtained from a recent version of BMR (Boix, M. Miller, and Rosato 2013). A dyad is a joint democracy if both were democracies according to the BMR indicator, if neither or only one was a democracy, the dyad is considered non-democratic. As mentioned previously, since we are specifically considering the causal effect of democratization, we need a binary indicator to identify joint-democratic dyads. Other covariates used were military

capacity of weakest dyad member, GDP per capita of the poorest dyad member, ratio of the two dyad members' CINC indices, and whether they were in a defense pact. We perform the analysis among politically relevant dyads, even though this subset of dyads do not capture the all militarized interstate disputes.

Whenever one of the democracies in a joint democracy dyad backslides, this dyad enters the control set. Dyads that become joint democracies again, after backsliding are recoded as new dyads and their baseline risk strata is re-updated based on the new transition time.

We choose a window of 20 years to classify dyads into their risk strata. Of the 1351 dyads that were a joint democracy at some point in the sample, 1210 dyads did not have a conflict in the 20 year look back window, and 141 did. The preponderance of joint-democracy dyads that have not been in a recent conflict is expected given that 70 percent of the 2420 dyads (included joint democratic and non-democratic dyads) have never been in a conflict.

Figure 4: The democratization effect conditional on baseline risk strata

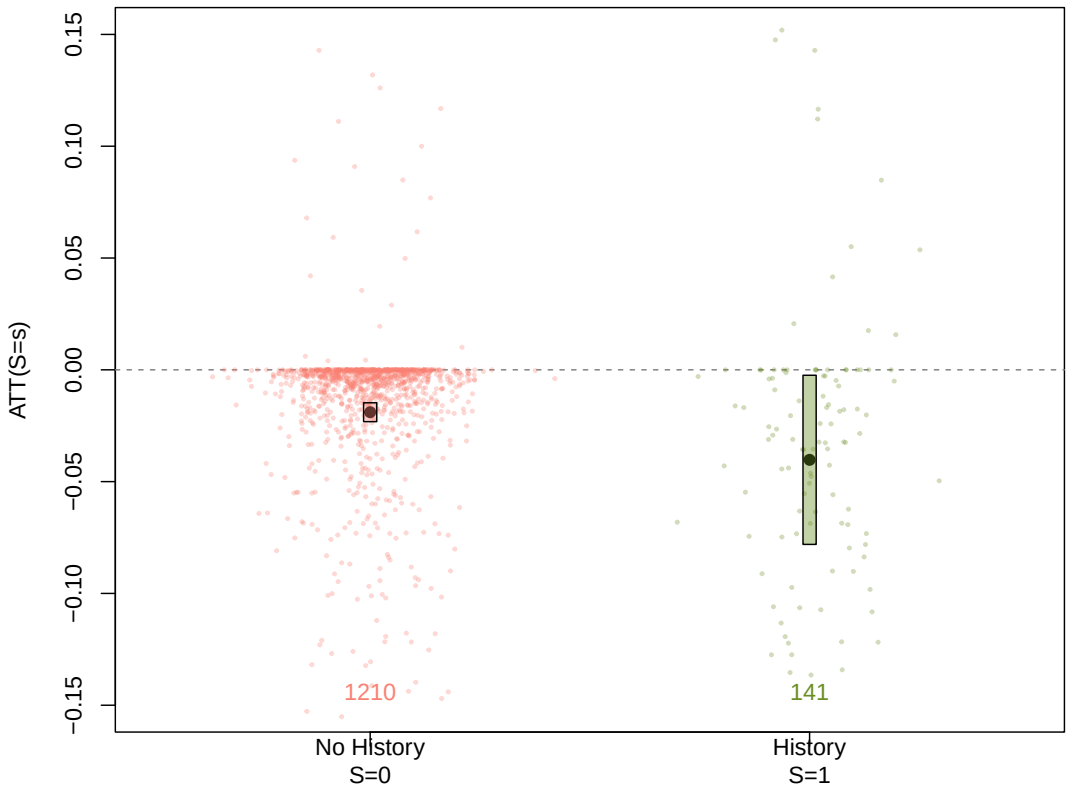


Figure 4 shows the democratic peace effect for the different risk strata: $\widehat{ATT}(S = 0)$

on the left and $\widehat{ATT}(S = 1)$ on the right. The post-treatment aggregates for each dyad, $ATT(i, s)$, are shown as the background dots. Under both strata, the democratic peace effect is affirmed. That is, conflict probability reduced after democratization. The effect is more pronounced in those that had a history of conflict (-0.04 vs -0.018). This pattern of results a similar to the predictions made in the previous section, although the magnitude of the difference in the treatment effects across the strata is smaller than in the simulated example (twice vs ten times). The overall effect, $oATT$, obtained by aggregating both these effects was -0.02. This is expected since there are far more joint democracy dyads with no history of conflict than there are joint democracy dyads with a history of conflict.

Figures 5 and 6 demonstrate the disaggregated $\widehat{ATT}(s, g)$ estimates and $\widehat{ATT}(s, t)$ estimates respectively. Even though we can estimate $\widehat{ATT}(s, g)$ for every year g, there are usually too few dyads that transition in every year. This is challenging for aggregation, so the reported $\widehat{ATT}(s, g)$ are for a coarsening of the transition time g, where all dyads that transitioned within the decade is aggregated to one. Even though the democratic peace effect was observed in both strata, figure 5 demonstrates that there is dyad-level heterogeneity in the effect. Among joint democracy dyads that had no history of fighting, the largest peace effect is for dyads that democratized in the decade from 1941-1950, after which almost all democracy effects were very close to zero. Among the dyads with history of fighting, the peace effect fluctuated rapidly in this same period from being negative in 1941-1950, to positive in 1971-1980, and back to being negative in 1991-2000.

Figure 6 demonstrates the effects $\widehat{ATT}(s, t)$. They demonstrate a similar pattern to the previous figure with dyads with no history showing little variation in the peace effect after 1960, and dyads with a history of conflict showing a lot more variation in the peace effect.

Figure 5: The democratization effect conditional on baseline risk strata and subset by democratizing decade

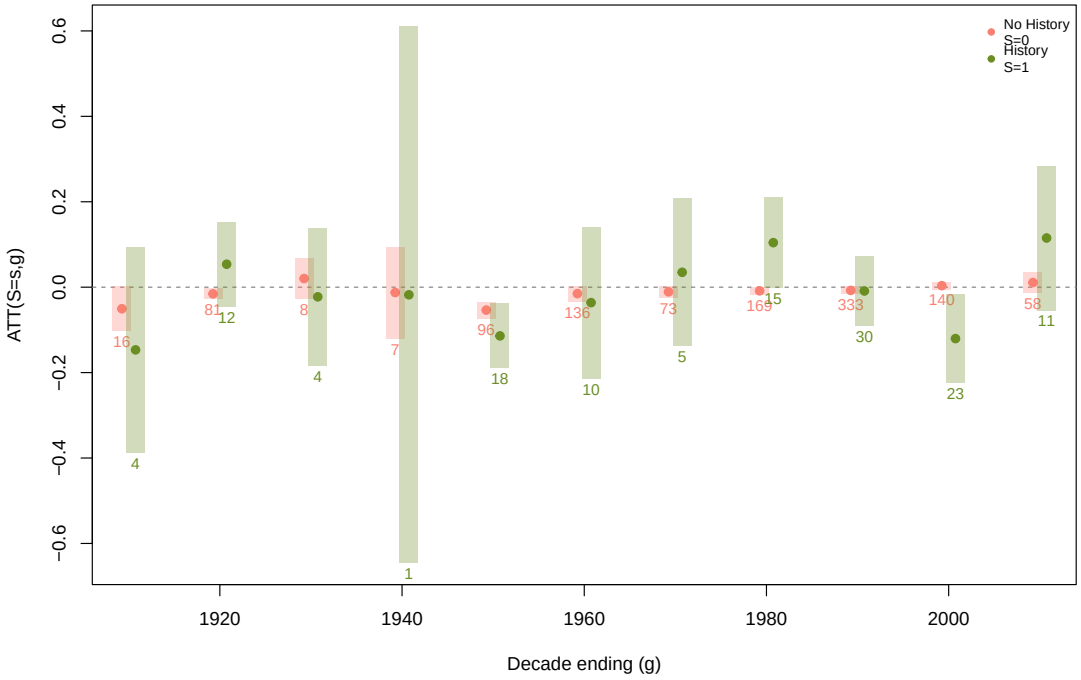
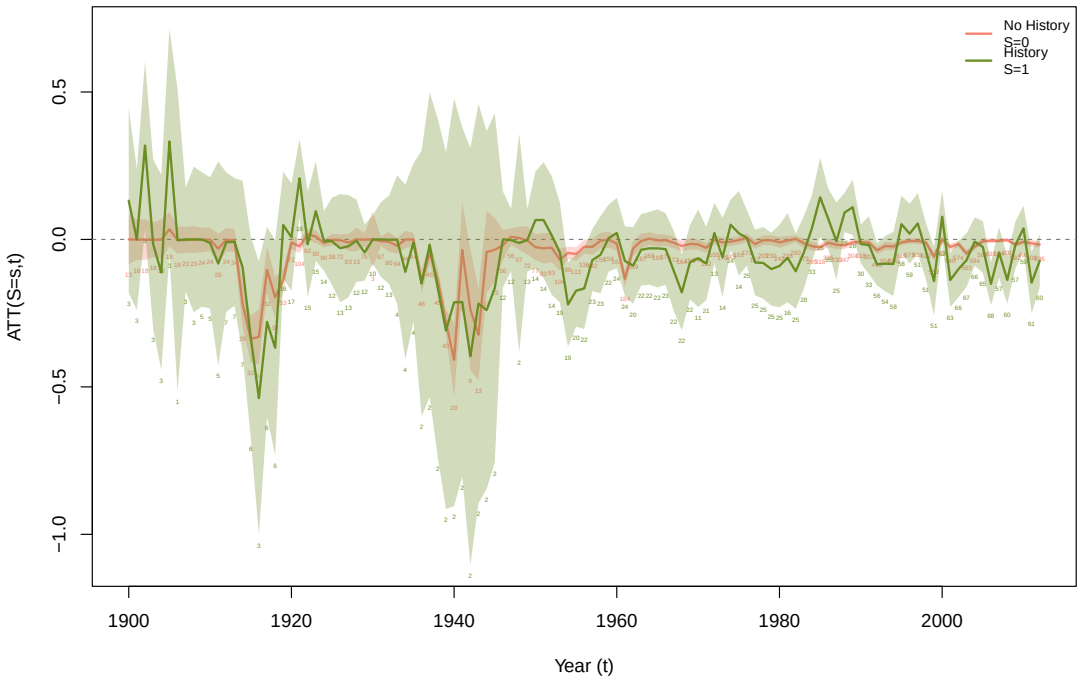


Figure 6: The democratization effect conditional on baseline risk strata and time



5 Conclusion

Many empirical studies testing the democratic peace theory estimate an average effect of demcratization. This paper points out that the focus on the average effect obscures the causal quantity that we are most interested in : the democratisation effect on dyads that have a high propensity to fight, or what we call belligerence.

Since most dyads, even among the politically relevant dyads, never engage in an interstate conflict, the average effect is severely depressed by the inclusion of these dyads. Another concern with an average effect is that there may be a selection effect depending on their hidden belligerence. Selection effects may be either that hidden belligerence facilitates democratization or hinders democratization, and may differ over time. Said another way, the mixing of belligerent dyads and non belligerent dyads exposes the average treatment effect to both attenuation and selection biases.

Our recommendation is to report average treatment effects conditional on a proxy for belligerence.

While there are many possibilities for such a proxy we use an encoding of whether the dyads has fought in a certain look back window. Using this risk proxy, when a dyad democratizes we force its control set to be other non-democratic dyads that also have the same risk encoding. This restricts dyads to have the same baseline (pre-democratic) risk. Other baseline differences between dyads are controlled for by inverse propensity score weighting. These estimates are the components that form the aggregates such as average treatment effect conditional on baseline risk. Aggregations can be even more granular than baseline risk.

Theoretical results from a model where hidden belligerence is a binary state show that splitting out average effects in this way inoculates against both the selection and attenuation biases. But it does not eliminate attenuation biases entirely since the risk proxy can misclassify non-belligerent dyads as belligerent as the window length increases.

Using a dataset on conflict data from 1820-2014, we show that the democratic peace effect for dyads with baseline risk of fighting is twice that of the dyads that had no baseline risk. However, dyads with baseline risk showed greater time heterogeneity in their effect, with those that democratized in the 1940s and 1990s being more peaceful and dyads that democratized in the 1980s being more conflict-prone.

References

Barber, Rina Foygel et al. (2021). "Predictive inference with the jackknife+". In: *The Annals of Statistics* 49.1, pp. 486–507.

Beck, Nathaniel and Jonathan N Katz (2001). "Throwing out the baby with the bath water: A comment on Green, Kim, and Yoon". In: *International Organization* 55.2, pp. 487–495.

Beck, Nathaniel, Jonathan N Katz, and Richard Tucker (1998). "Beyond ordinary logit: Taking time seriously in binary time-series-cross-section models". In: *American Journal of Political Science* 42.4, pp. 1260–88.

Boix, Carles, Michael Miller, and Sebastian Rosato (2013). "A complete data set of political regimes, 1800–2007". In: *Comparative political studies* 46.12, pp. 1523–1554.

Gartzke, Erik (2007). "The capitalist peace". In: *American journal of political science* 51.1, pp. 166–191.

Gibler, Douglas M and Steven V Miller (2024). "The militarized interstate events (MIE) dataset, 1816–2014". In: *Conflict Management and Peace Science* 41.4, pp. 463–481.

Green, Donald P, Soo Yeon Kim, and David H Yoon (2001). "Dirty Pool". In: *International Organization* 55.2, pp. 441–468.

Imai, Kosuke and James Lo (2021). "Robustness of empirical evidence for the democratic peace: A nonparametric sensitivity analysis". In: *International Organization* 75.3, pp. 901–919.