The “democratic peace”—the inference that democracies rarely fight each other—is one of the most important and empirically robust findings in international relations (IR). This article surveys the statistical challenges to the democratic peace and critically analyzes a prominent recent critique (Gartzke 2007). Gartzke’s claim that capitalist dynamics explain away the democratic peace relies on results problematically driven by (1) the censoring from the sample of observations containing certain communist countries or occurring before 1966, (2) the inclusion of regional controls, and (3) a misspecification of temporal controls. Analysis of these issues contributes to broader methodological debates and reveals novel characteristics of the democratic peace. Gartzke and other critics have contributed valuably to the study of IR; however, the democratic peace remains one of the most robust empirical associations in IR.

The “democratic peace”—the inference that democracies rarely fight each other—is one of the most important and empirically robust findings in international relations (IR). The apparent empirical association between joint democracy and peace has been debated and challenged since its first discovery by political scientists to the present (Gartzke 2007). Scholars have argued that this empirical association is in fact a product of other confounding factors, such as Cold War alliances (Farber and Gowa 1997; Gowa 1999), satisfaction with the regional status quo (Kacowicz 1995), shared foreign policy interests (Gartzke 1998, 2000), unmeasured factors such as dyad-specific effects (Green, Kim, and Yoon 2001), stable borders (Gibler 2007), and capital openness and development (Gartzke 2007; Gartzke and Hewitt 2010). Despite the large number of serious challenges, most current quantitative analyses continue to find a substantial, robust, and statistically significant association between joint democracy and the absence of militarized conflict. This article will analyze a recent challenge to the democratic peace (Gartzke 2007), situate it in the context of other statistical challenges to the democratic peace, and show that the democratic peace persists as a compelling finding. In so doing, this article also identifies new features of the democratic peace.

It is important to be clear about what this empirical association implies about international politics. Despite the robustness of this result to different model specifications, this observational finding by itself does not prove that it is characteristics of democracies—such as regular competitive elections, constraints on the executive, liberal norms, or civil rights—that make these countries more peaceful toward each other. Even less does it prove that the forceful spread of democracy in particular regions of the world will reduce the frequency or severity of wars. Justifying causal claims such as these exclusively using

---

Allan Dafoe is a Ph.D. candidate in the Travers Department of Political Science, 210 Barrows Hall, University of California, Berkeley, Berkeley, CA 94720 (dafoe@berkeley.edu).

I am deeply grateful to Giacomo Chiozza, David Freedman, and Bruce Russett for their time, insight, and encouragement. Thanks also to Devin Caughey, Chris Chambers-Ju, Sara Chatfield, Joanna Dafoe, Ron Hassner, John Henderson, Charlotte Jandér, Nina Kelsey, Matthew Kroenig, Rengyee Lee, Michael Mousseau, John Oneal, Ben Oppenheim, Ivo Pliek, George Willcoxson, Rick Wilson for editorial suggestions, and the anonymous reviewers. All errors are my own. Replication files and links to data are available at http://www.allandafoe.com/.

Additional analyses are reported in this article’s Supporting Information (SI).

1 A survey of IR specialists judged the study of the democratic peace to be the most productive research program in IR. Forty-five percent of the respondents of the TRIP Survey of International Relations judged “The Democratic Peace” to be one of the top three most productive controversies or research programs, scoring 9% higher than the closest runner-up (Maliniak et al. 2007).

2 Statistical associations are frequently referred to as “effects,” as in “B has a significant positive effect on C.” This terminology is convenient but misleading, as it implies, incorrectly in most cases, that conditional statistical associations are measures of causal effects.

3 “Joint democracy” refers to the fact that both members of a dyad are democratic.


©2011, Midwest Political Science Association

DOI: 10.1111/j.1540-5907.2010.00487.x
analyses of observational data requires the leverage of strong assumptions.

It is for this reason that there is less agreement about the actual causal mechanisms of the democratic peace than that around the underlying explanandum. Scholars have proposed that the democratic peace arises because of shared norms (Maoz and Russett 1993), restraint on democratic leaders (Bueno de Mesquita et al. 1999), more credible communication through transparency (Schultz 1998) or domestic audience costs (Fearon 1994; Tomz 2007; Weeks 2008), greater capacity to reach stable bargains (Lipson 2003), and other possible causal pathways. On the other hand, it may not be a “democratic” characteristic at all that accounts for the peace, but some other co-occurring or preceding factor, such as shared strategic circumstances, shared political systems, capitalism, prosperity, liberal economic norms, or other factors.

Nonetheless, the democracy-peace empirical association remains of paramount importance because, despite our best attempts to “control for” other possible correlates of this peace, the fact that two countries are democratic remains strongly associated with them having peaceful relations. Furthermore, under relatively modest assumptions this apparently peaceful proclivity seems unlikely to have arisen by chance (that is, the finding is “statistically significant”). This empirical association is foundational to a vast literature testing, refining, and extending theories about the apparent relationship between regime type and peace. Thus it matters greatly whether this association is robust to potential confounders (for reviews of this literature, see George and Bennett 2004; Ray 1995).

This article will analyze the claims made by an important and sophisticated critique of the democratic peace (Gartzke 2007). I will show that subtle issues in model specification problematically drive the results. Erik Gartzke argues that “economic development, capital market integration, and the compatibility of foreign policy preferences supplant the effect of democracy in standard statistical tests of the democratic peace” (2007, 166). However, Gartzke’s results against the democratic peace were not driven by the inclusion of variables for economic development or capital market integration, but rather primarily by (1) a censored sample due to the listwise deletion of those dyad-years for which the IMF measure of capital openness was unmeasured, (2) the inclusion of regional dummy variables, and (3) an incorrect implementation of Beck, Katz, and Tucker’s (1998) technique for controlling for temporal dependence. Analysis of these issues provides insight into the heterogeneous propensity of communist and noncommunist autocracies to have militarized disputes, temporal heterogeneity in the democratic peace, the ways that regional heterogeneity can be interpreted and addressed, the risks to inference of undertheorized control variables, and the distinctive patterns of temporal dependence in democratic dyads. In conclusion, this analysis finds that despite decades of attempted critiques, there remains a strong, statistically significant association between joint democracy and peace. Given the weight of evidence, readers should be properly skeptical of future claims to overturn such a thoroughly established result.

The Democratic Peace and Its Critics

A venerable lineage of scholars and thinkers have advanced arguments related to the democratic peace, including Georg Wilhelm Friedrich Hegel, Immanuel Kant, and Woodrow Wilson (Ray 1995, 4–6; Kant [1795] 1969). Scholars have proposed and tested a variety of propositions, such as whether democracies are less violent toward their own citizens (Rummel 1983) or against all other states (Rousseau et al. 1996). However, by far the most well-established proposition is that democracies are less likely to engage in violent conflict with each other. It is this apparent empirical association between democratic dyads and peace that makes up the foundation of the democratic peace research program. The dyadic democratic peace has faced many challenges, most of which can be categorized as either (1) raising concerns of potential confounding or (2) questioning the independence of the data and hence the statistical significance of results.

Concerns about the possibility of confounding have been present in the earliest works on the democratic peace. Scholars have sought to control for the possibility that the democratic peace is a spurious association driven by factors such as contiguity and distance, economic growth and wealth, the balance of military capabilities, alliance ties, and political stability. The most prominent criticism of confounding is that the democratic peace is a spurious association arising from underlying realist factors, such as alliance structures, the Cold War, or other shared foreign policy interests (Farber and Gowa 1995; Gowa 1999). However, the introduction of
control variables for all of these potential confounders into regression analyses has on the whole failed to remove the significant association between dyadic democracy and peace (Gartzke 1998, 2000; Kacowicz 1995; Lemke and Reed 1996; Maoz and Abdolali 1989; Maoz and Russett 1993; Oneal and Russett 1999b; Ray 1995; Rousseau et al. 1996; Russett 1993; Russett and Oneal 2001; Russett, Oneal, and Davis 1998; Signorino and Ritter 1999; Small and Singer 1976; Thompson and Tucker 1997).

A more recent lineage of challenges emphasizing omitted confounders, closely related to Gartzke’s (2007) critique, develops the argument that economic processes, such as the rise of commercial liberalism or increases in economic interdependence, are responsible for peaceful interstate relations, and that these factors correlate with or cause democracy. These studies have presented persuasive evidence that measures of economic factors strongly associate with conflict dynamics, and even account for some of the democratic peace. For example, Mousseau, Hegre, and Oneal (2003), building from Mousseau (2000), have shown that economically developed dyads are less likely to engage in militarized disputes, and that the democratic peace is conditional on the level of wealth in the dyad. More recently, Mousseau, using the per capita purchase of life insurance as a measure of the presence of a contract-intensive economy (CIE), makes the stronger claim that “economic development, [particularly] contract-intensive development, appears to account for [the democratic peace]” (2009, 53).

This stream of research has identified an important set of factors that condition the democracy-peace association; however, contrary to the stronger claims made, this literature does not as yet pose a fundamental statistical challenge to the democratic peace. Mousseau, Hegre, and Oneal (2003, 297) note that the interaction between regime type and wealth only renders the democracy-peace association insignificant for the poorest democratic countries. In their sample, this includes only 9% of the democratic dyads (Democracy > 6), and they note that by 1992 all democratic dyads are above this wealth threshold. Mousseau (2009) reports that the democratic peace is no longer significant when both dyads lack contract-intensive economies (oneCIE = 0). However, as Russett (2010) also points out, only 26% of the democratic dyads in Mousseau’s sample have no CIEs; for the other 74% of the democratic dyads Mousseau’s analyses find a significant association between democracy and peace. Thus, while Mousseau (2009) and related research have identified an important conditioning factor of the democratic peace, as with the realist critiques this set of work has not yet provided a compelling operationalization of a confounding factor that explains away the democratic peace.5

A second (related) class of criticism6 is potentially more severe, for it questions the statistical significance of the apparent empirical regularity by challenging the critical assumptions of independence in the typically used statistical tests. If conflict behavior for each dyad-year is not independent (conditional on the statistical model), then the confidence with which scholars reject the null hypothesis of no underlying association between joint democracy and peace may be exaggerated. Some scholars have questioned the significance of statistical results from analyses that do not account for temporal dependence (Spiro 1994); however, the democratic peace appears to be robust to various means of addressing temporal dependence (Beck, Katz, and Tucker 1998; Dafoe 2008; Russett 1995).

A more serious challenge arises in attempts to deal with cross-sectional dependence. In the context of information-poor datasets such as the binary rare-events international conflict data, common econometric solutions such as unit-specific intercepts (“fixed effects”) may not be practical since they do not leave enough information to reliably estimate the coefficients of slowly changing independent variables.7 Type II errors (failing to reject a false null hypothesis) are perceived to become excessive. The consensus seems to be that for fixed effects, the econometric cure is worse than the disease (Beck and Katz 2001; Green, Kim, and Yoon 2001; King 2001; Oneal and Russett 2001).8

Unlike temporal controls, cross-sectional controls have a substantial effect on the democratic peace.

---

5Mousseau also reports that pairs of countries both with CIEs (bothCIE = 1) never experience a fatal militarized conflict against each other. However, 76% of the countries coded as having a CIE also have the highest democracy score (Democracy = 10) and, for Mousseau’s data, pairs of countries with the highest democracy score (DemocracyLow = 10) also never experience a fatal militarized conflict against each other. Given the similarity between these empirical associations, greater caution is warranted before attributing this peace exclusively to CIE.

6A third class of criticism is that peace causes democracy (and not the reverse). Recent analyses do not find support for this position (see Crescenzi and Enterline 1999; Gates, Knutsen, and Moses 1996; James, Solberg, and Wolfson 1999; Maoz 2006; Midlarsky 1995; Mousseau and Shi 1999; Oneal and Russett 2000; Reiter 2001; and Thompson 1996).

7Though even with fixed effects, given a sufficiently long time series, the democratic peace is still apparent (Oneal and Russett 2001).

8Ward, Siverson, and Cao (2007) offer a critique of the “Kantian peace” (Oneal and Russett 1999c) based on modeling “higher-order dependencies.” For technical reasons they restrict their samples to single-year cross-sections. They report results consistent with the democratic peace.
Christopher Zorn (2001) points out that the cross-dyad relationship between democracy and peace is much stronger (four times) and much more significant ($p < 0.003$), as compared with the within-dyad association ($p \approx 0.4$). Similarly, scholars such as Kristian Skrede Gleditsch (2002) have shown how democracy and peace each exhibit strong spatial autocorrelation and seem to arise from spatial processes. Cross-sectional controls are thus likely to have a large effect on estimates of the democratic peace. Whether the inclusion of these controls will improve or worsen one’s inferences, however, depends on one’s theory, as will be discussed below.

In summary, statistical critiques of the democratic peace fall into two categories: challenges of confounding and challenges to the assumption of independence of observations. The former category of critique includes a variety of concerns and many important issues remain to be explored. Despite the many attempts to find confounding factors, however, the democratic peace remains a robust empirical association. The latter category of critique is to some extent a matter of epistemology, but conditional on a pragmatic burden of evidence standard to quantitative IR and a theoretically informed interpretation of results, this line of criticism has also failed to explain away the democracy-peace association. With reference to these categories of critique, I will now analyze in detail Gartzke (2007).

The Capitalist Peace

Erik Gartzke (2007), in an article that won the American Journal of Political Science’s annual award for best paper, shows that measures of capital openness, higher levels of economic development, and a measure of the similarity of state interests have a significant negative correlation with militarized interstate disputes (MIDs), and that the first two measures correlate significantly with fatal MIDs and war. Furthermore, Gartzke writes that his study “offers evidence suggesting that capitalism, and not democracy, leads to peace” (2007, 180). While his evidence of an association between his capitalist variables—capital market openness and economic development—and peaceful dyadic relations are noteworthy, well supported, and remain robust to altered model specifications, it is not those variables that weaken the statistical association between democracy and peace in his analyses.

Gartzke’s basic statistical model involves estimating the following logistic regression, where $DV_{ij,t}$ refers to the respective dichotomous dependent variable, either the presence of a MID, a fatal MID, or a war, and $p_{ij,t} = Pr(DV_{ij,t} = 1)$:

$$\ln \left( \frac{P_{ij,t}}{1 - P_{ij,t}} \right) = \beta_0 + \beta_1 DemocracyLow_{ij,t-1}$$

$$+ \beta_2 DemocracyHigh_{ij,t-1} + \beta_3 Trade\ DependenceLow_{ij,t-1}$$

$$+ \beta_4 Contiguity_{ij} + \beta_5 LogDistance_{ij,t-1}$$

$$+ \beta_6 MajorPower_{ij} + \beta_7 Alliance_{ij,t-1} + \beta_8 LogCapabilityRatio_{ij,t-1}$$

$$+ \beta_9 Africa_{ij} + \beta_{10} Asia_{ij} + \beta_{11} Europe_{ij}$$

$$+ \beta_{12} MiddleEast_{ij} + \beta_{13} NorthAmerica_{ij}$$

$$+ \beta_{14} SouthAmerica_{ij} + \beta_{15} Spline_{1ij,t-1}$$

$$+ \beta_{16} Spline_{2ij,t-1} + \beta_{17} Spline_{3ij,t-1}$$

There is one observation for every year (from 1950 to 1992) per pair of countries (dyad-year). Huber-White standard errors, clustered by dyad, are computed. The independent variables are mostly those used in Oneal and Russett’s work (see, for example, 1999a).  

11War is defined as a MID with over 1,000 fatalities.

12For more detail on the construction of the variables, see (Gartzke 2007, 174–76). Democracy combines the democracy and autocracy scales from Gurr Polity IV Data (Jaggers and Gurr 1995) into a measure from $-10$ to 10. DemocracyLow$_{ij} = \min (Democracy_{ij,1}, Democracy_{ij,2})$, DemocracyHigh$_{ij} = \max (Democracy_{ij,1}, Democracy_{ij,2})$, TradeDependenceLow is the lower dyadic measure of the ratio of bilateral trade over GDP. Contiguity indicates whether the two countries in the dyad share a land border or are separated by less than 150 miles of water. LogDistance is the natural logarithm of the distance between two states’ capital or closest major cities. MajorPower indicates whether at least one member of the dyad is China, France, the United States, the United Kingdom, or the USSR. Alliance indicates comembership in a defense or neutrality pact (Small and Singer 1990). LogCapabilityRatio is the natural logarithm of the ratio of the stronger state’s Correlates of War capabilities index to that of the lower (Singer et al. 1972). Region names indicate whether both members of a dyad are within
FIGURE 1  Replication of Gartzke 2007

This figure illustrates one of the key results of Gartzke (2007): that the estimated association between DemocracyLow and Peace becomes small and insignificant after capitalist variables are added (compare model G4 and G1 in the dotted rectangle). The difference in predicted probability of a MID from one SD change in an independent variable (x-axis) is calculated using models G1 (\(N = 282,287\)) and G4 (\(N = 171,509\)) for a dyad with a prior probability of a MID of 0.5. The independent variables are those listed on the y-axis, plus Temporal Splines (without peace years) and a constant. See SI for the table of results and descriptive statistics. *Variables with an asterisk are dummy variables. For these, the change in predicted probability is also based on a one SD change so as to make results more comparable (a one-unit change can readily be calculated using the fact that, where \(q\) is the mean of a dummy, one standard deviation of a dummy \(\sqrt{q(1-q)}\)).

Figure 1\(^{13}\) illustrates Gartzke's central finding about the apparent effect of capitalist variables on the democracy-peace correlation. Model G1 refers to Gartzke’s basic model (Model 1), meant to replicate Oneal and Russett (1999a), in which both dyadic variables for democracy have a large and statistically significant association. Model G4 refers to Model 4 in Gartzke (2007) in which capitalist variables are added. DemocracyLow is the lower dyadic measure of the 21-level Polity measure of democracy. FinancialOpennessLow is the lower dyadic measure of a nine-level indicator compiled by Gartzke, Li, and Boehmer (2001; see also Gartzke and Li 2003) from an IMF assessment of “government restrictions on foreign exchange, current, and capital accounts” (Gartzke 2007, 174). GDPLow is the lower dyadic measure of GDP per capita, and GDPcontiguity is an interaction term between GDPLow and Contiguity. Gartzke’s results were precisely

\(^{13}\)I follow the recommendations of Gelman and Hill (2007, 551) and others (Kastellec and Leoni 2007) in using figures, rather than tables, to communicate key comparisons.
replicated, thanks to his making his data and replication files available. Gartzke’s claim that the democratic peace is explained away by the inclusion of capitalist variables rests on the change in magnitude and statistical significance of the estimated coefficient of DemocracyLow in G4 and similar “capitalist” models (see dotted rectangle for key comparison).

In order to argue that it is the inclusion of his capitalist variables that drove his finding against the democratic peace, Gartzke chose to adopt the model setup of the democratic peace research program as exemplified by Oneal and Russett (1999a) because it “allows for ready comparison of results and diminishes the danger that [his] findings result from idiosyncrasies in coding or model specification” (Gartzke 2007, 173). In at least three respects that are not discussed, however, Gartzke’s models deviate from Oneal and Russett’s specification; these differences generated the evidence against the democratic peace. After accounting for these three specification decisions, the inclusion of capitalist variables has no discernible effect on the association between DemocracyLow and peace (in Figure 5, this zero effect is revealed in contrasting “Models 0” and “Models G”; the result presented by Gartzke contrasts “Models A” and “Models G”).

**Missing Data and the IMF Sample Bias**

Gartzke’s findings against the democratic peace are driven largely by the listwise deletion of observations with missing values of the IMF measure of financial openness. This can be seen in Figure 2. Row “G1 on Full Sample” presents a replication from Gartzke’s baseline model of the estimated association of DemocracyLow and DemocracyHigh with peace; both low and high dyadic values of democracy have significant and large estimated coefficients. The next row (“G1 on G4 Sample”) presents the results from running the same regression on the censored sample associated with G4, and thus only on those observations for which there exist values for financial openness and GDP per capita. This censoring of 39% of the dataset makes the association of both dyadic measures of democracy weak and statistically indistinguishable from zero. This contrast makes it apparent that most of the change in the estimated coefficient and significance of DemocracyLow occurred because of the change in sample, and not from the inclusion of the capitalist variables per se. Finally, the third row (“G1 on Complement of G4 Sample”) presents the results from running the model specification from G1 on the complement of the sample for G4, that is, only for those dyad-years for which values of financial openness or GDP are missing. On this smaller sample, both dyadic measures of democracy have large estimated coefficients and are highly significant.

In sum, there seems to be substantial heterogeneity in the relationship between democracy and conflict when pairs of countries in the IMF dataset are compared with pairs of countries not included in the IMF dataset. In fact, effectively all of the change in the democracy-peace association generated by Gartzke’s capitalist models (contrast Models A and Models G in Figure 5) can be attributed to the effect of censoring the sample.

Gartzke’s Model 1 (G1, also equivalent to Models A) is run on available data for all countries from 1950 to 1992, but the IMF measure of capital openness does not begin until 1966 (leading to missing values for 59,442 dyad-years). After that, 48,333 dyad-years are excluded because of missing values on capital openness and 3,485 are excluded because of missing values on GDP (with 442 dyad-years overlapping). Seventy-eight percent of the dyad-years with missing data after 1965 involve at least one communist country. The remaining dyad-years do not have anything apparent in common.14

Figure 3 illustrates the differences between the IMF and the non-IMF samples. Specifically, Figure 3 graphs the proportion of dyad-years experiencing a MID, broken up by DemocracyLow (x-axis) and whether the dyad-year is in the IMF sample (colored columns). The primary difference between the IMF and non-IMF samples (compare black and light grey columns) occurs when DemocracyLow is −9, −8, and −7 (though also in the −4 to 2 range). China, the Soviet Union, Vietnam, and North Korea seem to be driving the different conflict behavior for these categories of autocracy (as noted on the figure).

I implement a technique, described in the SI (see particularly Table 3), to isolate the effect on the estimated DemocracyLow coefficient from adding particular countries to the G4 sample in the period after 1965. I found that the greatest effect came from the inclusion of North Korea, Vietnam, China, Angola, and the Soviet Union. The measure of development—real GDP per capita—does not have missing values for many countries, and those countries with the most missing values (Niger and Barbados) do not substantially change the DemocracyLow coefficient. In summary: the democratic peace is notably stronger in samples that include the years 1950–65 and that include communist countries.

---

14The main noncommunist countries that have missing values for this measure of capital openness are Bhutan, Brunei, Luxembourg, Maldives, Switzerland, Taiwan, and Zimbabwe.
This figure illustrates how the estimated association between DemocracyLow and peace (and DemocracyHigh and peace) is weak when using only the IMF sample, and otherwise large and significant. Predicted probabilities are calculated using model G1 under different samples, with a prior probability of a MID of 0.5. The “G4 Sample” refers to the set of observations that have values for FinancialOpennessLow and GDP; this excludes dyad-years before 1966 and most dyads with a communist country.

Following Gartzke’s argument, I briefly assess whether the belligerent tendency of communist autocracies could be better attributed to their closed capital markets, rather than their nondemocratic character. I impute values of financial openness to the countries excluded by the IMF sample in a set of ways generous to the hypothesis that levels of financial openness account for the conflict behavior formerly attributed to regime type. Under seven different imputation rules, using Gartzke’s capitalist models (Models G) with a corrected temporal specification (as will be discussed below), DemocracyLow is always significant at least at the 0.001 level with respect to MIDs and the 0.1 level with respect to wars. Correlations with fatal MIDs are less clear, though the estimated coefficient is always significant once regional dummies are removed (discussed below). Thus, variables for financial openness and development, even when their missing values were imputed in ways likely to reduce the statistical effect of regime type, do not “supplant the effect of democracy in standard statistical tests of the democratic peace” (Gartzke 2007, 166).

Missing data is a serious problem in many social science datasets, and prominent methodologists have cautioned against listwise deletion (King et al. 2001). Reaffirming this recommendation, the above section showed how Gartzke’s results against the democratic
This figure illustrates how the association between DemocracyLow and peace (columns get shorter moving right) is conditional on membership in the IMF sample. The proportion of dyad-years (left axis) that experiences a MID is illustrated, broken up by categories of DemocracyLow (x-axis) and membership in the IMF sample (different-colored columns). The line graphs the number of dyad-years (right axis) per category of DemocracyLow.

peace were driven not by the inclusion of capitalist confounders but by the censoring of dyad-years—those prior to 1966 or involving communist countries—in which the democratic peace is particularly strong.

Cross-Sectional Dependence

The second element of Gartzke’s model specification that was different from Oneal and Russett’s and which undermined the correlation between joint democracy and peace was the inclusion of regional dummy variables. The exclusion of regional dummies from Gartzke’s analyses makes the estimated coefficient on DemocracyLow much stronger and statistically significant in two out of three of his principal models (see Table 5 in SI; also contrast Models 0 and Models B in Figure 5). To give a sense of the magnitude of the effect of including regional dummies, the estimated effect of a one s.d. increase in DemocracyLow on the probability of conflict is more than double when regional dummies are excluded (the reduction in the probability of conflict from a one s.d. increase in DemocracyLow increases from 0.015 to 0.044, from 0.038 to 0.084, and from −0.052 to 0.077 for MIDs, fatal MIDs, and wars, respectively, for a dyad with a 0.5 baseline probability).

Gartzke’s justification for this deviation from Oneal and Russett’s work is that “several scholars identify regional variability in interstate conflict (Bennett and Stam 1999; Lemke 2002, 2003a, 2003b). Controlling for sample heterogeneity is important on both econometric and substantive grounds [particularly] in the context of the democratic peace (Henderson 2002)” (2007, 176). The inclusion of regional dummies, however, is problematic in this case.

The most significant problem with Gartzke’s inclusion of regional dummies is that they have no apparent relationship to his theoretical arguments about the effect of capitalist variables, but they had the potential to generate results different from Oneal and Russett’s work. Gartzke notes that it is important to adopt Oneal and Russett’s statistical model because it “allows for ready comparison of results and diminishes the danger that [the] findings result from idiosyncrasies in coding or model specification” (Gartzke 2007, 173). However, regional dummy variables have not been used by Oneal and Russett, and their inclusion substantially weakened the democracy-peace correlation for reasons unrelated to the capitalist peace.

Some might argue that the estimates from models with regional dummies should be taken seriously, and
thus that Gartzke (2007) inadvertently reveals the sensitivity of the democratic peace to regional control variables. This raises the question about under what circumstances it might be appropriate to add some form of control for cross-sectional heterogeneity, such as dummy variables for regions, countries, leaders, or the most significant subset of these. One reason that scholars may want to use cross-sectional controls is as a simple means to test the robustness of a particular result to potentially omitted cross-sectional factors. However, the implications of a nonrobust association depend on the particular statistical model and controls and the theoretical framework motivating them. This section will argue that a set of cross-sectional controls is only likely to improve one’s inference if one believes that the cross-sectional variation is reasonably modeled using the specific variables and functional form of the cross-sectional controls. This requires that (1) the candidate cross-sectional controls are unlikely to proxy for consequences of the independent variables of interest, and (2) the controls are sufficiently theoretically motivated so as to support the requisite claim that they are correctly specified. Regional dummy variables do not seem to meet these criteria.

The use of regional dummy variables implies that regional variation is reasonably modeled as differing baseline levels of conflict across the regions specified. However, the lack of theoretical justification for regional dummies could as readily give rise to a different specification. Consider an alternative specification based on the conjectures that (1) the most democratic (autocratic) countries may exert a regional pacifying (exacerbating) effect on interstate conflict relations, and (2) that regional dummies correlate with spatial proximity and will thus absorb some of the imperfectly specified association between distance and conflict (analyses are described in detail in the SI).

To operationalize the first conjecture, I created two dummy variables, RegionalDemocracy and RegionalAutocracy, that identify whether in any particular spatial area there exists a country with the highest or lowest combined polity score. RegionalDemocracy is an annual variable that equals 1 if both members of any dyad are within 1,000 miles of each other and are each fewer than 1,000 miles away from a country with the highest polity score, 0 otherwise. I selected 1,000 miles because that is approximately the radius of regions such as Europe and the Middle East. RegionalAutocracy is similarly constructed, but using instead the lowest polity score.

The second conjecture builds from the intuition that the relationship between distance and conflict may be imperfectly approximated by the logarithmic functional form of the prevailing control for distance, leading statistical models to underestimate the probability of conflict for those dyads that are close to each other (fewer than 3,000 miles, and especially fewer than 500) and overestimate the probability of conflict for those dyads that are more distant from each other. To partly correct for this probable functional misspecification, I substitute three variables—distance, distance², and distance³—for the logarithm of the distance. See Table 6 in the SI and the replication files for details.

These two new sets of variables were progressively introduced to Model G4, and in the last model regional dummies were removed. Each of these models was progressively preferred by BIC (a measure of fit that penalizes for model complexity). The inclusion of these two sets of variables to G4 makes all regional dummies insignificant except Europe (because most regions merely proxy for proximal dyads), and the regression reveals that the two new regional regime-type variables are important and highly significant (the presence of either is associated with a change in the predicted probability of a MID greater than 6%, p < 0.01). With the removal of regional dummies, a moderate statistically significant association is estimated for the standard regime-type variables (one s.d. increase in DemocracyLow is associated with a 5% reduction in the predicted probability of a MID, p < 0.10). Thus, the heterogeneity meant to be addressed by undertheorized control variables may be better addressed by other specifications, with superior fit and theoretical justification, that lead to different results and theoretical implications.

Most proposals to control for cross-sectional heterogeneity explicitly or implicitly make the necessary theoretical claim that the cross-sectional controls proxy for exogenous causal factors and do not proxy for other endogenous factors. For example, fixed effects could be justified in terms of unmeasured geographical features, such as the Himalayas (Green et al. 2001, 142), which are clearly causally prior to other variables. Likewise, regional dummies could be used to proxy for plausibly exogenous

16 A number of scholars have similarly expressed skepticism about the value of undertheorized control variables. For statements in international relations, see Achen 2005; Clarke 2005; and Ray 2003, 2005.

17 The results I report are robust to moderate modification of this length.

18 Standardized Pearson residuals, among other diagnostics, reveal this pattern.

19 Technically, the distance between their capitals or major cities for large countries.
factors such as geography, climate, or disease conditions. Justifications such as these are desirable because they provide an opportunity to debate the causal role of the factor in question, and they direct efforts to refine the operationalization of the variable so as to be as close to the concept as possible.

Undertheorized controls, on the other hand, evade such criticisms of theory and measurement, and consequently risk serving as proxies for endogenous factors. “Europe,” for example, could proxy for the postwar liberal bargain in Europe that arose from the victory of democratic forces (Judt 2005, 63–100). By this interpretation, then, a large and statistically significant coefficient on “Europe” could be read as evidence in favor of a particular interpretation of the democratic peace. Similarly, dyadic dummy variables may proxy for long-standing unresolved rivalries, which are known to be much more common among nondemocratic dyads (Diehl and Goertz 2001, 101–27). Statistical results can only be interpreted through a theoretical lens. Results based on undertheorized variables do not lend themselves to a clear interpretation and can be used to support multiple theoretical claims.

Zorn’s (2001) finding that the cross-sectional evidence for the democratic peace is much stronger than the longitudinal evidence, along with the modest sensitivity of the democratic peace to regional dummies, suggests a direction for future research. Most theories of the democratic peace propose a contemporaneous dyadic mechanism—such as greater utility from peace (due to norms or domestic politics) or more efficient negotiation (through transparency or domestic audience costs)—and therefore anticipate a longitudinal as well as cross-sectional association. The relative weakness of the longitudinal evidence therefore poses a challenge to these theories in a manner similar to how the absence of a longitudinal association between development and democracy calls into question many prevailing theories of democratization (Robinson 2006). Dyadic contemporaneous theories of the democratic peace thus need to be modified. Perhaps the pacifying effect of dyadic democracy is slow to take effect? Perhaps it is conditioned on the political neighborhood, such as through regional pacific externalities? Perhaps the “third wave of democratization” (Huntington 1992) was less pacific than the first two, which largely occurred before Zorn and Gartzke’s datasets begin (in 1950)?

The cross-sectional character of the democratic peace may pose a fundamental challenge to many current theories. This variation needs to be theorized and explored, however, not “controlled.” If a primary objective of social science is “getting rid of proper nouns” (King 2001, 504; see also Przeworski and Teune 1970)—that is, to replace ideographic descriptions with generalizations and causal theories—then introducing controls such as “Europe” into one’s regressions should be done with caution. They may efficiently improve the fit of one’s model at the cost of controlling away or obscuring that which needs to be explained (Gleditsch 2002, 11). While robustness to the inclusion of undertheorized controls affirms the robustness of an association, the sensitivity of a result to undertheorized controls is not compelling evidence against the presumed association. 20

Controlling for Temporal Heterogeneity

The third component of Gartzke’s specification that inadvertently weakened the correlation between democracy and peace was a subtle error in his implementation of Beck, Katz, and Tucker’s (1998) technique of adding temporal splines to control for temporal dependence. Technically, the technique proposed by Beck, Katz, and Tucker (BKT) for adding temporal splines to one’s regressions requires that a “peace-year” variable be included along with the “spline” variables generated by their BTSCS tool for Stata (Beck, Katz, and Tucker 1998). Gartzke used BKT’s technique but without including a “peace-years” variable, thus generating an estimate of the hazard rate with a poor fit to the data and also generating a weaker estimate of the coefficient of DemocracyLow. 22 Though in general the concerns discussed above about cross-sectional controls may also apply to temporal controls, since it was Gartzke’s intention to implement Beck, Katz, and Tucker’s technique, this section will focus on the statistical consequences of the misimplementation of this technique.

20Choi (2011) also points out the sensitivity of Gartzke’s (2007) results to the exclusion of region dummies (and temporal controls). Choi argues against the inclusion of region dummies because they led to the dropping of observations in some analyses (due to perfect prediction in Gartzke’s analysis of wars). Dropping perfectly predicted observations, however, is reasonable if one is confident in one’s statistical model; it is equivalent to estimating a coefficient to be infinitely large. The issue with region dummies is not that observations are dropped but that they are generally inappropriate for analysis of the democratic peace.

21The splines for this analysis were generated with three knots and, as is commonly done by scholars of international relations, using the cumulative number of years of dyadic peace as the time unit. See the replication files for more details. For a cautionary discussion of temporal controls, see Dafoe 2008.

22Oneal and Russett (1999a) used an alternative technique for implementing splines.
Figure 4 Estimated Odds-Ratio Over Time for Gartzke and Corrected Specifications (MID)

This figure illustrates how the estimated hazard (expressed in odds-ratios, left axis) from Gartzke’s temporal specification (dashed line) does not fit well with the actual pattern of MIDs per peace-year (bar graph: the proportion is on the right axis, and the relative risk—which approximates the odds-ratio—can be read off the left axis). The estimated hazard from a corrected temporal specification (solid line) fits much better. Model G1 minus regional dummies was used.

While altering the temporal specification by adding the peace-years variable to Gartzke’s regressions leaves most substantial results the same, it does make the estimated coefficient on DemocracyLow stronger and more statistically significant, and greatly improves the model’s fit.24 Figure 6 in the SI illustrates why a misestimation of the hazard has this consequence for the estimation of the DemocracyLow coefficient. In summary, democratic dyads, as compared with mixed or autocratic dyads, are less likely to have MIDs that spill over across multiple years and are less likely to have new MIDs occur in the immediate years following a MID, relative to their long-run risk. This may reflect the tendency for democratic dyads to avoid becoming rivals.

Democracy Still Matters: Three Consequential Specification Decisions

Gartzke’s statistical evidence against the democratic peace arose from three problematic features of his model

23Gartzke’s dependent variable included every year of a MID. Counting only the onset of MIDs substantially reduces the hazard at the 0th peace-year.

24The same temporal specification issue occurs in Gartzke, Li, and Boehmer (2001), though without major effects.
specification: the listwise deletion of dyad-years with missing values for capital openness, the inclusion of regional dummy variables, and an error in the implementation of temporal controls. Figure 5 summarizes how changing these three components of his model affects estimates of the democratic peace (also see Table 7 in SI).

Models G are Gartzke’s final “capitalist” models (precisely replicated) with the dependent variable as the presence of a MID (Model 4), fatal MID (Model 9) or war (Model 7). The rows show how the estimated association between DemocracyLow and peace changes subject to the alteration of the three specification decisions. Models 0 are Models G minus his capitalist variables, but still using the censored sample from Models G. Models A, B, and C then each start with Model 0 and modify one of the three critical specification decisions. Models A expand the sample to include the non-IMF observations and corresponds to the baseline models in Gartzke (2007). Models B remove the regional dummy variables. Models C correct the temporal specification. The rest of the models then combine these modifications. Models A + B + C are thus close to the standard statistical tests familiar in Oneal and Russett’s work.

A few things are worth noting from this figure: (1) Holding the sample constant, variables for capital openness and development (contrast Models G and 0) make virtually no difference to the estimated association between DemocracyLow and peace. (2) Gartzke’s finding of an insignificant association between DemocracyLow and peace is fragile: it only arises for MIDs if all three specification issues are present, and for fatal MIDs and wars if the sample is censored and at least one other issue is present. (3) The estimated democracy-peace association is substantial. For Models A+B+C, a one s.d. increase in DemocracyLow is associated with a 13%, 10%, and 26% reduction in the predicted probability of MIDs, fatal MIDs, and war, respectively.

This last point can be emphasized by comparing the estimated associations for other independent variables. Using Model G4, but with corrected temporal...
Specification, no regional dummies, and values of 0 imputed to FinancialOpennessLow for communist countries after 1965, a one s.d. increase in DemocracyLow is estimated to reduce the probability of a MID by 10% ($p < 0.001$). A similar one s.d. increase in the log of the distance of the capital cities, in the lower-dyadic capital openness, or in the lower-dyadic per capita GDP of a contiguous dyad reduces the estimated probability of a MID by, respectively, 8%, 5%, and 5%. There are certainly issues with comparing the estimated associations from standardized changes in variables (King and Zeng 2007), but it provides a means of comparing estimated associations from “comparable” changes in the independent variables. Using Gartzke’s adjusted data and model, therefore, the above analysis shows that even when controlling for capital variables, democracy has a strong and significant association with peace, seemingly stronger than other covariates such as distance, capital openness, and wealth.\footnote{In a recent article, Gartzke and Hewitt (2010) investigate the onset of ICB crises using a model specification similar to Gartzke (2007). They partly address two of the concerns raised in this article: they employ a corrected temporal specification, and they impute values of 0 for FinancialOpennessLow for dyads that are missing values after 1965. Russett (2010) reports, however, that as in Gartzke (2007), the inclusion of FinancialOpennessLow leads to a critical reduction in the sample (dropping dyad-years before 1966). A dummy variable for the Middle East also contributes to the reduced DemocracyLow coefficient.}

**Conclusion**

The democratic peace is one of the most thoroughly established empirical relationships in international relations. Deserving of its prominence, the democratic peace has met with many critiques. Scholars have charged that the association is spurious, driven by some confounding factor such as realist interests or capitalism, and that the association could in fact arise by chance given the degree of dependence in the data. Despite the many critiques, however, the democratic peace remains a substantively important and statistically unlikely association by high standards of inference for the study of international relations. The recent critique by Gartzke (2007) similarly fails to provide compelling evidence against this finding, as his results were driven not by capitalist variables but by a series of problematic specification decisions.

Gartzke (2007) shares themes with the earlier critiques of the democratic peace. From these themes a number of specific and general conclusions about the statistical study of the democratic peace, and observational data more generally, can be drawn. First, new variables that correlate with democracy have been frequently introduced as potential candidate confounders; variables for capital openness, distance, alliances, wealth, and others may proxy for important factors in international relations. They have not, to date, removed the underlying association between joint democracy and peace. Second, the inclusion of undertheorized variables like regional dummies is motivated by the awareness of potential dependence in the data. Undertheorized controls provide useful checks on the robustness of results. However, in the absence of theoretical justification and construction, descriptive and causal inferences are muddied by their inclusion. Third, statistical analyses, even when performed by sophisticated and careful scholars, risk being sensitive to subtle aspects of model specification; for this reason scholars should set themselves the high standard of making complete replication materials readily available on publicly accessible websites.

A more general point emerges about the role of descriptive inference in causal inference. The democratic peace is foremost a descriptive inference (Berk 2004, 203–18; King, Keohane, and Verba 1994, 34–74): it is a statement about an empirical association that is robust to conditioning on a large set of covariates and statistical models. This empirical association is often used to support the causal inference that it is a particular property of democracy that reduces the risk of conflict in democratic dyads. The persuasiveness of this argument depends on one’s theories and the likelihood of the empirical association given these theories about the international system. However, there are many plausible theories that posit different causal relationships between such factors as liberalism, democracy, capitalism, and development. For example, Schumpeter argued that “modern democracy is a product of the capitalist process” (1943, 297); Olson argued that “the conditions necessary for a lasting democracy are the same necessary for the security of property and contract rights that generates economic growth” (1993, 567); Przeworski et al. (1996) proposed that development promotes the survival of democracy; Robinson (2006, 523–24), finding little credible evidence that income per capita promotes democracy, theorizes that development and democracy are by-products of a country’s historically determined development path; North, Wallis, and Weingast identify an institutional affinity in history that they call the “double balance: open access and entry to organizations in the economy support open access in politics, and open access and entry in politics support open access in the economy” (2009, 24; see also Mousseau 2009).

Thus, there are many plausible causal connections between capitalism, income per capita, and democracy.
Given weak priors over the causal models relating these factors, the descriptive inference of the democratic peace does not rule out the possibility that it may be capital-
ism, development, or other processes that are the direct, and/or indirect, cause of the liberal peace. Nor, importantly, would a robust statistical finding that an economic covariate absorbs the association between joint democracy and peace be sufficient evidence that democracy is not a direct, let alone indirect, cause of peace. Strong theory or a credible source of exogenous variation is required to draw clear causal inferences from observational data.

This being said, the resilience of the democratic peace to numerous plausible confounders is important support for the view that democracy is an essential component of the liberal peace. Future scholarship will continue to challenge, refine, and theorize about this association, and future critiques are likely to continue on the trajectories outlined in this article. Scientific progress, however, seems less likely to come from continued statistical “horse-races” between intertwined macrohistorical factors than from research that leverages credible identification strategies (Angrist and Pischke 2010) or that specifies and tests the mechanisms purportedly generating these macroassociations (e.g., Tomz and Weeks 2010). As the number of studies supporting the descriptive inference of the democratic peace continues to grow, the probability of a future study overturning this finding becomes increasingly less likely. Scholars should weight the informativeness of new evidence accordingly.

References


