"Dirty Pool" Revisited: When Less is More

Robert S. Erikson
Political Science Department
Columbia University
420 W 118th Street
New York, NY 10027
212-854-0036
rse14@columbia.edu

Pablo M. Pinto
Political Science Department
Columbia University
212-854-3351
pp2162@columbia.edu

Kelly T. Rader
Political Science Department
Columbia University
202-413-3179
ktr2102@columbia.edu

Prepared for delivery at the 2009 American Political Science Association Convention, Sept.3-6, Toronto, Canada

Abstract

Among IR scholars, a central empirical proposition is that democracies seek out other democracies as trading partners—the so-called democratic trade hypothesis. Yet, as revealed in a 2001 symposium on Green et al.'s "Dirty Pool" testing this hypothesis is entangled in debates over the appropriate statistical techniques and research design. We use this controversy as a springboard to offer a cautionary tale about how a large data set with a massive N can create overconfidence in hypothesis testing. On the one hand we have over 90,000 dyads of nation-years. On the other hand we can observe only a small number of national transitions in and out of democratic status. These considerations suggest that the proper estimation of a democracy effect (and its standard error) are not readily solved by mechanical resort to statistical formula, particularly with dyads as the units of analysis. Our central contribution is to employ randomization tests on the dyadic analysis to infer the correct *p*-values associated with the main hypotheses. Second, we model nation-years, where the question is whether the proportion of trade with other democracies increases when a country becomes more democratic. Third, we conduct a difference-in-difference analysis of change in trading partners following democratic or anti-democratic shocks. Finally, we embed our nation-state results in a multi-level framework, distinguishing between the short-term effects of democratic transitions on trade from the long-term effects of national democratic (or not) culture on trade. Rather than adding further layers of statistical complexity, these tests actually are simple and quite intuitive.

I. Introduction

Do democracies trade more with other democracies? It is widely believed that they do. But do democracies actually seek out other democracies with which to trade? To find out, numerous quantitative studies have been conducted with dyads composed of trade partners as the units of analysis. These studies have typically found supportive evidence.[1] But there have been dissenting voices.[2] As one would expect, the answer to the substantive question of democracy and trade gets tangled up with important methodological questions about the appropriate statistics and research design with which to investigate the question.[3]

In 2001, Green, Kim, and Yoon (2001) [hereafter GKY] published an influential but controversial article in *International Organization* that challenges the conventional wisdom regarding both that democracies are more likely to trade with one another and that democracies are more peaceful with each other than non-democracies or mixed dyads. Regarding both hypotheses, the central question was whether or not fixed "dyad" and "year" effects should be incorporated into the analysis, with the substantive results clearly hinging on the answer. While the controversy regarding a democratic peace may be the more important of the two issues, the dichotomous measure of war and peace also makes it the more daunting question to analyze statistically. In the present paper, we address only the lesser but seemingly more tractable question of democratic trade.

GKY's trade data includes information on more than 90,000 dyad-years involving as many as 115 countries over 30 years.[4] Following the established convention in the literature for how the question should be posed, the question was: does the democracy score of the least democratic of the two trading partners affect (presumably positively)

---

[1] Morrow, Siverson, and Tabares (1998, 1999) and Bliss and Russett (1998) were among the first scholars to analyze the effect of joint democracy on trade. See also Mansfield, Milner, and Rosendorff (2000).

[2] See Garrett (2000); O'Rourke & Taylor (2006); Dai (2002). The main argument against the democratic trade hypothesis is grounded on the expected distributive consequences of trade as mediated by political institutions. Trade is likely to affects the wellbeing of different actors in the polity in different directions depending on their ownership of factors of production or sector of employment. See Stolper and Samuelson (1941), Jones (1965), Samuelson (1965). Whether democracy leads to demand for protection depends on the ability of the free trade or protectionist lobbies to affect policy-making, which is likely to covary with political institutions (Mayer 1984, Rogowski 1989, Alt and Gilligan 1994, Hiscox 2002).

[3] The methodological approach uses dyad-year as the unit of analysis following the empirical literature on the "democratic peace" in IR. See Doyle (1986); Maoz and Russett (1993); Bremer (1993); Oneal, Oneal, Maoz, and Russett (1996); Oneal and Russett (1997).

[4] The number of countries varies over time since GKY restrict the analysis to dyads for which twenty or more observations were available (Green et al. 2001, 451).

the level of trade between the two nations composing the dyad?[5]

GKY show that the answer to the question whether democracy promotes trade between democracies depends on a crucial methodological decision: whether to include in the regression analysis "fixed" effects for the dyad units. Incorporating fixed effects is standard procedure in econometric analysis, but apparently novel to International Relations research at the time. GKY first conducted the analysis according to the usual convention with typical controls based on the "gravity" model of international trade and both with and without a lagged dependent variable. This conventional OLS analysis produced the usual result found in the literature—a highly significant effect for the democracy score for the dyad's least democratic member.[6] Then GKY added fixed effects—over two thousand dummy variables (!), one for each dyad in the analysis. This innovation upended the result. The sizeable positive democracy effect vanished, with or without controlling for lagged effects. In fact, the sign of the democracy coefficient flipped to negative. The first four models in our Table 1.1 display GKY's findings as reported in Green et al. (2001).[7]

Given GKY's counterintuitive findings, *International Organiz*ation invited other prominent methodologists to enter the fray. Beck and Katz (2001) articulated the view that while fixed effects are often appropriate their application to this problem was erroneous, especially with controls for lagged effects. Oneal and Russett (2001) argue that fixed effects were inappropriate, and that besides, if one were to use fixed effects there should be not only fixed unit (dyad) effects but also fixed time (year) effects as well. Adding year controls to the fixed effect analysis, Oneal and Russett was able to reverse the sign of the coefficient.[8] King (2001), generally sided with the critics rather than GYK, with most of his analysis focused on the democratic peace rather than democratic trade.

According to the critics, controlling for fixed effects is over-control—"throwing the baby out with the bathwater." To get the results that theory says make sense, they approve using OLS without fixed effects. All symposium contributors expressed some degree of uncertainty about the value of the information from so many dyads over so many years, with discomfort over whether the inclusion of 90,000 plus cases offered a false sense of

[5] The question could be posed in different terms by looking at specific examples: would the volume of trade between the US and China be larger if China was democratic?

[6] This is the conventional measure of joint democracy in the literature. See Green et al. (2001), Bliss and Russett (1998), and Gowa and Mansfield (1993).

[7] Table 1.2 reproduces the GKY results after transforming bilateral trade to constant 1996 dollars.

[8] The fifth column on Table 1.1 reports the results obtained when fitting a model with fixed country and year effects and a lagged dependent variable to the GKY dataset. Contrary to Oneal and Russett (2001) we find that the coefficient on minimum democracy for the dyad turns negative albeit not significant. Note, however, that the results reported by Oneal and Russett are for a longer time span covering the period 1886-1992, not the postwar era.

precision.

II. Our initial take

Numerous methodological issues are left unsettled by the 2001 symposium. First of all, it must be startling to the lay reader to find that with a whopping 90,000 cases there remains uncertainty about the sign of the democracy effect and the statistical significance of its estimate. How can so much depend on a seemingly arcane matter of methodology—should one include fixed effects (and perhaps which ones) or not? Why did not the symposium contributors come to grips with the substantive implications of the results varying so much by model specification? Was somebody very wrong, or do we have the grounds for being sure of our estimates?

Actually, there is a standard diagnosis when the introduction of fixed effects alters the estimated effect of the treatment variable on the dependent variable. GKY's null finding with fixed effects means that averaged over thousands of dyads, there is no tendency for the dyad members to trade with each other more in years when the least democratic member is at its most democratic. In the face of this finding, the seemingly positive result when fixed effects are omitted signifies that democracy and trade must be related in the cross-section. Thus, we have a seeming contradiction between the within-dyad and between-dyad trade pattern. On the one hand, even when controlling for other variables in the model, democratic dyads engage in the more intra-dyad trade than non-democratic dyads. Yet dyads do not shift their within-dyad trade level when they undergo democratic change.

As we see it, GKY are correct on the crucial question of whether fixed effects should be employed. There is no doubt that the Hausman test regarding whether fixed effects can be ignored leads to a rejection of that idea. Because the key coefficients vary with the presence of unit (nation) effects, the inclusion of unit effects is required. Given the huge *N*, the loss of a few degrees of freedom is trivial. This is standard econometrics. Further, this applies both to the static model and the dynamic model (with lagged effects of the dependent variable). By excluding fixed nation effects from the dynamic model (Beck and Katz's proposed solution), the statistical implication is that (controlling for the independent variables) at equilibrium each dyad reverts to the identical level of trade.

Still, the fixed effects results may be relevant only regarding the short-term and not so relevant to the long-term selection of trading partners. The time-series and cross-section effects represent related but different phenomena. From the equations in Table 1.1 (and 1.2) we might infer the following. When a nation changes its status on the democracy ladder, such as experiencing a coup or deposing its ruling junta, the change does not induce a large short-term disruption of its relationship with its trading partners.[9] While

[9] The data suggests that the reluctance to date of the US and other democratic governments to apply economic sanctions to Honduras after the ousting of President Manuel Zelaya on June 28th 2009 is not an exceptional event.

there is a possibility that a government (or group of governments) could choose to sanction a trading partner for violating democratic practices, the general dispositions of a set of two nations to trade to with each other (the fixed dyad effects) is in part a function of the mutual affinity of their cultural traditions which includes their long-term tendency regarding the level of democracy of their institutions.

Apart from proper model specification, estimation of the standard error and statistical significance of the democracy coefficient is a particular challenge. Although we have upwards from 90,000 cases in the form of dyad-years, should each dyad-year be treated as an independent case? For instance, if a dyad displays a constant democracy score for 30 years, does it present 30 independent cases (as it does in the pooled analysis), or simply one? Even more problematic and specific to the dyad as the unit, consider the example of the initially nondemocratic nation *A* that shifts toward democracy in year *t*. This transition gets recorded in all 100+ dyads involving nation *A* for year *t*. Should all 100+ dyads be treated as individual units, or as one event?

III. Dyads as the problem

While we side with GYK on the use of fixed effects, that is not the central question to the present paper. Our critique goes to the matter of using dyads as the unit of analysis. At a minimum, they present unrealistic standard errors which yield overconfidence in the results. We suggest that the question of democratic trade is better analyzed by abandoning the dyad in favor of the nation as the cross-sectional unit and also imposing a multi-level model. That is, time-serially, the question should be: do countries trade more with other democracies when they themselves are at their most democratic? And cross-sectionally, conditional on short-term effects and year-adjusted, are nations that generally trade more with other nations more democratic themselves?

For computing the standard error of the democracy coefficient, we need the appropriate degrees of freedom. With *N* nations, the number of undirected dyads (*D*) is

$$\frac{N!}{(N-2)!2!} = \frac{N(N-1)}{2}$$

Is the effective number of degrees of freedom the *D* x *T* (dyads times years) total, yielding over 90,000? [The pooled case.] Is it (D x *T*) – D -1? [The fixed effects case] Or should the effective number of cases be the *N* x *T*, the number of nation-years? Or is the relevant N the small number of observable democratic transitions? These are crucial issues. Standard errors are deflated in proportion to the assumed number of degrees of freedom. If we base our analysis on an erroneously large interpretation of the degrees of freedom, the danger is to become too overconfident—perhaps much too overconfident—in our results.

The present paper offers a fresh approach to the question of democracy and trade, using a similar data set to GKY's 90,000+ case data set described above. First, our central contribution is to employ randomization tests to infer the correct *p*-values associated with the main hypotheses in the democratic trade debate. Where the correct assumptions about the variance of the estimators (e. g., variance of OLS coefficients) are unclear, the nonparametric procedure of randomization tests is a useful alternative. Second, we offer our own model, where the cross-sectional unit is the nation and not the dyad. The question becomes whether democracies comprise more of a nation's trading partners when the nation is at its most democratic. The units are nation-years rather than dyad years, and fixed nation (and year) effects are employed. Third, we analyze a selected set of pro-democratic and anti-democratic regime changes and conduct a difference-in-difference analysis to see whether change in democratic trade follows regime change. Fourth, we return to the nation-state data and embed it in a multi-level framework, distinguishing between the short-term effects of democratic transitions on trade from the long-term effects of national democratic (or not) culture on trade. Here we measure a nation's democratic trading and its democracy score as residuals from the time-series analysis. Rather than adding further layers of statistical complexity, these tests actually are simple and quite intuitive.

IV. Time and Inflation

First, we deal with a mild complication. As Oneal and Russett point out, GKY measure trade in terms of nominal dollars whereas it is far more defensible to measure trade in real dollars that would control for inflation. Time enters another way as well. Trade and democracy may both be changing over time in tandem due to factors independent of any causal effect of democracy on trade such as common shocks. If so, the results are distorted. The solution to that problem is to employ fixed time effects (with a set of dummy variables for each year) in addition to the nation (unit) fixed effects.

Accordingly, as our first step after replicating GKY's equations (the first four equations in Table 1.1) is to replicate the same tables with real (1996) dollars instead of nominal dollars and (for fixed effects equations) add year dummies along with the dyad dummies. Results with the additions of years (but still, nominal income) to the dyads in the fixed effects model are shown in equations 5 and 6. The addition of year dummies causes the oddly negative coefficients for democracy to plunge by more than half. Without dynamics, there is little change in the tiny reported standard error, so the negative coefficient still appears quite significant with fixed year effects. With dynamics, the democracy coefficient drops sufficiently to lose its rating as statistically significant.

The variations in the original models when trade is measured in constant 1996 dollars are shown in Table 1.2.[10] In no instance does the estimate of the democracy effect change much compared to its sister equation from Table 1.1 where trade is measured in nominal dollars. Again, the democracy effect looks alive and well with pooling, whether or not dynamics are included. Again, the democracy effect has the "wrong" (negative) sign

[10] The variables and sources are discussed in Green et al. (2001, pp. 450). We thank Donald Green for sharing the data with use.

with fixed dyad effects. In sum, our exercise replicating GKY with deflated trade data suggest that the estimated coefficient on the lower regime score in the dyad is sizable when pooling, but turns negative and not significant when adding dyad and/or time dummies.

V. Randomization tests

As we argue above, it is difficult to know how to properly calculate the standard errors on (and thus test the significance of) coefficients obtained from dyad-year data because it is unclear what the effective number of observations is. The standard *t*-tests used in GKY rely on the classical assumption that the disturbances in the model are independently and identically distributed with mean zero and standard error σ^2. In other words, there is no unaccounted for heterogeneity in the data, such as clustering at the dyad or country level or autocorrelation over time. This assumption seems implausible to us, but what the correct model of the disturbances should be is also elusive.

Instead, we propose a randomization test that does not rely on any assumptions about the shape of the disturbances in the data. This non-parametric technique is a takeoff on Fisher's (1935) exact test, and it proceeds in a way that is analogous to a parametric hypothesis test. (For a modern overview, see Edgington and Onghena 2007). Basically, this test compares the observed coefficient to the density of false coefficients obtained when the national identities are scrambled in a series of simulations. By this test, we find that the estimates of the democracy effect are within the realm expected by chance from random assignment of country democracy profiles.

When conducting a hypothesis test, we want to know how likely it is that an estimated coefficient is different from the null hypothesis, usually that the coefficient equals zero. To perform a *t*-test like those employed by GKY, we first estimate standard errors with the assumption that disturbances are distributed i.i.d. $N(0, \sigma^2)$. We then derive a *t*-test statistic by taking the ratio of the estimated coefficient and the estimated standard error and compare that statistic to a student's *t* distribution. If it is larger than the critical value 1.96 or smaller than -1.96 (for sample sizes 1000 or larger), we reject the null hypothesis of no effect at the 95 percent confidence level.

Randomization tests require a similar procedure but do not rely on a theoretical distribution like the student's *t* distribution, and thus their validity does not hinge on assumptions about the disturbances in the data. First, we estimate the model coefficients and their associated test statistics in the typical way, using the OLS and fixed effects models described in Table 2. Then, we randomly reshuffle the data in order to break the systematic relationship between the democracy score and the observed trade level.

To do so, we first scramble the nation labels for democracy scores, as if we drop the nation labels on the floor and reinsert the nation-labels randomly. For instance, a simulation might find the dyad Ecuador-Belgium is given the democracy score for Japan-Costa Rica, while the Japan-Costa Rica dyad is given the democracy scores of Norway-

Egypt, and so on. For each simulation, each dyad is assigned the democracy scores for the same pair of nations for each year the dyad appears in the analysis. Moreover, for each simulation, each nation is assigned the democracy scores for one unique nation for every dyad it is involved in. While cases are scrambled for nation, they are not scrambled for year. Thus, in the example above, Ecuador would be assigned Japan's democracy scores for all dyads for all years and Belgium would be assigned Costa Rica's democracy scores for all dyads and all years. After the democracy scores are randomized, the minimum democracy score within the dyad, the key independent variable, is recalculated.[11]

Then, we re-run the models on the shuffled data and get a new estimate of the coefficient and test statistic, knowing that the true coefficient and test statistic should be zero on average. We reshuffle and re-estimate a total of 1000 times. For each of the simulations, Ecuador and Belgium, for example, will be assigned still different democracy identities.

This process gives us a distribution of 1000 estimated coefficients and 1000 test statistics centered, theoretically, at zero. These empirical distributions, instead of theoretically derived distributions, are the reference distributions for the randomization test. We can calculate a non-parametric *p*-value by locating the observed effect (the estimated test statistic) on this distribution and measuring what proportion of the 1000 test statistics are larger in absolute value than the absolute value of the observed test statistic. And so we are able to properly assess the probability that democracy coefficient could have occurred by "chance."

We can use the randomization test results as both a method for correctly conducting hypothesis tests for a given dataset without resorting to parametric assumptions and as a robustness check against the conventional *t*-tests. If the randomization *p*-values on the democracy scores are similar in size to the *p*-values that are reported from the actual data, then our concerns about the effective sample size in dyad-year data will be incorrect. If, however, the randomization *p*-values are larger, then we can conclude that the conventional test is inappropriate for this dataset.[12]

VI. The Randomization Test Data Set

For the randomization test, it would be ideal to use the same dyad-year data employed by GKY. However, the unbalanced nature of the GKY data makes this impossible. The GKY dataset is unbalanced due to missing data for the dependent variable and other

---

[11] Setting aside the extra step to recalculate the dyad-level democracy score, this approach is similar to a "block randomization" technique (Manly 1997). In social science applications, it is analogous to the "placebo law" method (Helland and Tabarrok (2004), Donohue and Wolfers (2006)).

[12] In general, parametric techniques are valid only if they generate p-values that are close in size to those from a randomization test (Moore, et al 2003, p57, Edgington and Onghena 2007, p.289).

covariates, and changes in the number of countries in the world caused by state birth and death. Thus, in our randomization test simulations, we cannot randomly assign the 1950 democracy score for, say, Mozambique, to Canada (or the reverse) because Mozambique did not exist in 1950.

To improve the coverage for the dependent variable we use the expanded bilateral trade dataset created by Kristian Gleditsch, and we are able to obtain more complete data on other covariates. From this expanded dataset, which at times has as many as 143 countries, we are able to identify a rectangular matrix of 2346 undirected dyads comprising 69 countries with complete coverage for the 1950-2000 period that could be used for our randomization tests.

Table 2 reestimates the equations from Table 1.2, now using our balanced data set. For the pooled model, the estimated democracy effects remain highly significant. For the fixed effects models, the democracy coefficients have changed. The democracy scores coefficients for the least democratic dyad member, that with the original data were mysteriously negative, now, mysteriously, have turned positive. The relevant models are those with dyad fixed effects alone and those with both year and dyad fixed effects. Without dynamics (equations 2 and 5), this democracy coefficient is both positive and significant. With dynamics, the positive coefficient is trivial and but still statistically significant.

At this juncture, the reader is excused by being made dizzy by the topsy-turvy findings. Let us focus on the question of whether dyads are appropriate. Even if we pretend that dyads are the appropriate unit, the source of the variance in findings could be that the reported variance estimates for the coefficients are simply wrong. In other words, apart from the very real specification issues, should we believe the standard errors? To find out, we turn to the randomization analysis in the following section.

VII. Randomization Test Results

In order to reassess the confidence with which one can state findings based on dyad-year level analyses, we subjected the democracy coefficients in the models in Table 2 to a randomization test. The *p*-values obtained from our randomization tests are shown below the conventional parametric *p*-values on the democracy coefficients in Table 2. Again, they are calculated by asking what proportion of the 1000 *t*-test statistics calculated in the random democracy score shuffles is larger in absolute value than the absolute value of the observed *t*-test statistic from the unshuffled data.

First, as we suspected, the randomization test *p*-values are substantially larger than the parametric *p*-values. In each of the seven models, the randomization test *p*-values are at least *45 trillion* times larger than the conventional *p*-values. Thus, conventional significance tests dramatically overstate the confidence of results obtained with dyad-year data, even after accounting for different dyad intercepts (as in equations 2 and 4), different year intercepts (as in equation 6), or both (as in equations 5 and 7).

Because the conventional *p*-values are too small, they are highly likely to lead a researcher to make a type I error—that is, to conclude that there is a significant democracy effect when indeed there is none. In each of the 1000 randomization test runs, we recorded the *p*-values associated with the minimum democracy score coefficients calculated using the conventional significance test on the scrambled data. Because the time series of minimum democracy score was randomly assigned, we know that it has no meaningful association with trade within the dyad. Thus, if the conventional test is appropriate for this data, we should (incorrectly) reject the null hypothesis only 5 percent of the time at the 95 percent confidence level, 10 percent of the time at the 90 percent confidence level, and so on.

Figure 1.1 shows the distributions of the 1000 conventional p-values calculated during the randomization tests for each of the models in Table 2. The shaded areas cover *p*-values that are 0.1 or smaller, small enough to justify rejecting the null hypothesis that minimum democracy score within the dyad has no effect on trade. Even though these *p*-values were calculated using data in which we know no systematic relationship exists, between 60 percent and 94 percent of the *p*-values were less than or equal to 0.1. In other words, using conventional significance tests this dyadic data would cause one to falsely infer that the meaningless minimum democracy score within the dyad has a significant effect on trade 60 to 94 percent of the time (depending on the specification), instead of 10 percent of the time, as we would expect at the 90 percent confidence level. Again, this extreme over-confidence is consistent across all of the specifications, even those that allow for varying dyad and year intercepts.

Even though the conventional tests have unacceptably high type I error rates, that does not necessarily mean that there is no significant democracy effect. Indeed, despite the marked overconfidence of the standard hypothesis tests, the coefficients on the minimum democracy score within the dyad remain statistically significant at conventional levels. In the pooled model and the pooled model with dynamics, democracy effects are still statistically significant according to the randomization tests, with randomization *p*-values 0.001 and 0.014 respectively. In the models with dyad fixed effects only, the coefficient on the minimum democracy score with the dyad is significant, with a *p*-value of 0.011 without dynamics and 0.013 with dynamics. In the model with year fixed effects alone, the minimum democracy coefficient is significant with a *p*-value of 0.001.

However, as we argue above, the appropriate specification for dyad-year data is one that includes both dyad and year fixed effects. When both dyad and year fixed effects are included, the coefficient on the minimum democracy score within the dyad, though substantively small, is still significant following the randomization test, both without dynamics (p=0.036) and including dynamics (p=0.022).

Figure 1.2 presents the randomization test results graphically. The graphs on the left-hand side show the densities of the 1000 coefficients on the democracy variables estimated using the 1000 datasets in which each country's democracy score over time is randomly assigned. The dark gray shaded areas represent the 5 percent most extreme coefficients, and the entire shaded areas cover the 10 percent most extreme coefficients. The dotted lines indicate the magnitudes of the coefficients estimated using the actual observed data, the coefficients in Table 2. From these graphs, we can see that the estimated minimum democracy effect is sufficiently large to be rare by conventional statistical standards in each of the models.

The graphs on the right-hand side show the densities of the 1000 *t*-statistics calculated using the conventional *t*-tests of significance in the OLS and fixed effects models during each randomization run. As in the coefficient graphs, we see that the test statistics derived from the actual data are among the 5 percent most extreme for each of the models, including our preferred specification.

The importance of including time fixed effects is particularly evident in these graphs. Recall that, in theory, the distribution of coefficients and *t*-statistics obtained during the randomization should be centered at zero. In other words, when the minimum democracy score within the dyad over time is scrambled, there should be no systematic effect between it and the level of trade on average. However, in the pooled and pooled with dynamics models, the distributions are centered at negative values. This means that, *even when the democracy score times series are random noise*, one finds a negative relationship on average between democracy and trade in data with this structure. In the second and fourth models, those with dyad fixed effects alone, the distributions are centered at positive values. This means that, *even when the democracy score times series are random noise*, one finds a positive relationship on average between democracy and trade. Because democracy and trade both increase on average over time, almost any random arrangement of democracy score will show a positive coefficient, controlling for across-dyad variation. Only when time fixed effects are included, as in the last three models, are the distributions centered at zero as they should be.

To summarize, the randomization results show that conventional *t*-tests are inappropriate for testing hypotheses on dyad-year data because they rely on highly overconfident standard errors. That the randomization test *p*-values are orders of magnitude larger than the typical *p*-values means that this dataset does not meet the assumptions required for the reporting of conventional standard errors and *t*-tests. However, using correctly specified dyad and year fixed effects models and the appropriately non-parametric randomization test, we can still reject the null hypothesis of no effect of minimum democracy score within the dyad on trade, albeit at a much lower confidence level.

VIII. How Democratic is your trade? Using Nations as the Units

The democratic trade hypothesis is about nations, not dyads.[13] The hypothesis is that when a nation increases (decreases) its level of democracy, it gains (loses) democratic

[13] Mansfield, Milner and Rosendorff (2000) are an exception. Yet their model is about trade negotiations the joint determination of tariffs through bilateral bargaining, not volumes.

trading partners. For the nation under a microscope, the presumed mechanisms are its shifting motivation to trade with democracies plus other nations' shifts in taste for trading with it.[14] Thus the relevant unit is the nation rather than the set of dyads to which it is connected in the analyst's data set. A nation's economic leaders may have a set motivation to trade with democracies in general (or reciprocally leaders of other democracies may prefer trading with it); but this is one decision rather than multiple unconnected decisions regarding trading with each of the countries current and potential partners. And as we have seen, treating each dyad-year as an independent unit of observation yields false confidence in the estimated effects of democracy on trade.

In this section we use nation-years as the unit of analysis. We attempt to explain the nation-year's percent of trade that goes to democratic nations, or simply its score on "democratic trade." More specifically, democratic trade is measured as the nation's percentage of its trade in year *t* that is with other nations that hold a score of six or greater on the -10 to +10 democracy scale.[15] The independent variable of interest is the nation's democracy score in year *t* on the same democracy scale, a number from -10 to +10. Our analysis covers all of the 6297 nation-years for the period 1950-2000 for which the democracy score and bilateral trade is measured. We also use the smaller set of 3519 nation-years comprising the 69 x 51 balanced panel of nation and year data that comprise the nations of our dyadic analysis.

The idea for the basic equation is simple: a regression where the Y_{it} variable is the percent of trade with democracies by nation *i* in year *t* and the X_{it} variable is nation *i*'s democracy score in year *t*. Then we raise the same issues as before. Fixed effects or not? Do we include a lagged dependent variable to account for time dependence? A further model, and an important one, that we consider is a fixed effects model assuming an AR1 autoregressive process in order to account for autocorrelated error.

Table 3 shows the results. In each equation, the basic independent variable is X_{it}, the nation's democracy score. The only other right-hand side variables are possible fixed effects for nation and year.

The table shows that with complete pooling (a simple bivariate model), each nation's rate of trade with democracies increases roughly one percent for each extra peg it climbs on

[14] Preferential treatment granted to specific trading partners could result from preferences at the governmental (helping other nations stay democratic, or sanction those that deviate) or individual levels (consumers discriminating against goods and services produced in countries that violate civil and political liberties usually associated with democratic practices).

[15] The percent of trade with democratic nations is based on other nations in our dataset for which democracy scores are measured. Like previous studies we use the (arbitrary) cutpoint of +6 on the Polity IV composite score for a country to classified as democratic. The 6+ cutpoint for consolidated democracies was originally proposed by Farber and Gowa (1997), and is the value recommended by Marshall, Jaggers and Gurr (2007). See also Polacheck (2007, pp. 1053).

the 20-point scale of democracy. But of course this is wrong. We turn to a panel design, presenting random and fixed effects. With random effects and with fixed effects for nations but not years, the coefficient declines slightly. Adding fixed effects for years as well as nations drops the coefficients down further, to 0.38 for the full data set and 0.63 for the balanced panel, which contains a greater proportion of economically developed set of nations than the full data set.

The drop in coefficients with the year dummies is a function of the growth of democracy throughout the half century. As nations are becoming more democratic over time, both the independent variable and the dependent variable must grow. Thus, to ignore time dummies is to proclaim results as correct which are in large part a statistical artifact.

The low democracy coefficients of 0.38 and 0.63 in the equations with nation and time fixed effects may seem low. They imply that the effect of going from the most undemocratic nation to the most democratic (the full 21 point range) range is only an increase of 7 to 13 percent in the share of trade that is exchanged with democratic nations. Still, the large N assures a highly significant (if small) effect. Figure 2 illustrates for the full data set, showing the residualized rate of democratic trade as a function of the residualized democracy score (each controlling for nation and year). The relationship is visible, but clearly nonzero only by the power of over six thousand cases.

So it might seem that the democratic trade hypothesis must be true after all. But there is at least one more problem. The within-nation prediction errors are highly autocorrelated. The point is that apart from a nation's level of democracy and time effects, its democratic trade score will not be randomly distributed over its time series. Rather they will cluster temporally so that nations have higher than normal trade scores at some time intervals and lower than normal scores at others.[16]

The standard econometric solution for this problem is the Prais-Winston method which assumes that the errors evolve as an AR1 process (so that at each time point it is a function of only the error at the preceding time measurement). The AR1 fixed (nation, year) effect model shows a startling result. Not only do the coefficients drop further in magnitude; now with thousands of cases, statistical significance becomes a challenge. With over 6217 cases, the estimate for the full data set is significant (barely). But the coefficient for the 3159 cases in the balanced panel no longer is statistically significant. The culprit is the high autoregressive coefficient (rho) in the 0.90 range. Even the OLS fixed effect result with nation and year controls appears to be an illusion.

IX. Democratic and anti-democratic shocks

Consider the problem. For most nations, democracy scores vary little (if at all) over

[16] The existence of serial correlation in the data cannot be rejected. The *F*-stat on the Wooldridge (2002) test for autocorrelation in panels is 214.095 with a p >.000, which rejects the null of no first-order autocorrelation in the errors (See Drukker 2003).

time. For most nations, democratic trade scores rarely change, as reflected in the correlated error structure. The best evidence consists of nations that maybe once or twice make some sort of shift on the democratic scale. The analyst's slim reward is the rare nation with a sufficiently unstable political history that it provides at most one or two instances of meaningful democratic change. These shifts are the events that drive the results. There are not 90,000 of them. There are not 3,000 of them. With the detailed answer depending on the definition of what magnitude of shift is large enough to count as a democratic (or anti-democratic) shock, the answer is less than 50 cases. In our analysis below, the number of shocks is a mere 38 cases. As we will see, this modest collection of case study material provides useful statistical evidence regarding the democratic trade hypothesis.

We identify 26 democratic shocks and 12 anti-democratic shocks. To meet these criteria,

- A nation's democracy score undergoes a one-year shift of at least six points.
- The shift crosses the +6 threshold, thought by some to be the best cutoff value for the presence of democratic institutions as discussed above.
- There must be five available observations on trade and democracy for five years before the shift and five years after, thus creating an 11-year interval of observations.
- There must not be another shift of magnitude |6| or greater during the 11-year interval that would present a complicating overlap.

The yield is 26 instances of democratic shocks and 12 instances of anti-democratic shocks. On average, these were not small shocks. The average pro-democracy gain is 11.2 points on the 21-point scale. The average anti-democracy loss is a plunge of 13 points. Our case selection consists of real instances of democratic change. The question now is, did these nations change their trading patterns to trade more (less) with democracies as a response to the shock?

Figures 3 and 4 show the picture for gains in democracy. Figure 3 represents trade as raw scores. Figure 4 provides some focus by measuring each trade score as a deviation from its mean over the 11 years presented. If we assume a linear model for years on this scale predicting democratic trade, the coefficient is 1.71 percent increase for every year during the eleven-year period of democratic transition. This coefficient is highly significant whether or not the observations are nation-adjusted. Adjusted, in effect the units are merely the eleven time points, yet the t-statistic achieves a lofty 10.78 reading. Unadjusted, with an N of 286 (11 x 26), the t-statistic is 6.57.

In fact the democratic trade vs. time relationship shown in Figures 3 and 4 is not linear—there is more gain in democratic trade in the run-up to the democratic shock than after. And the trend is running with the current, since on average over our full data set of 6,000+ cases the gain in democratic trade per year is 0.43 percentage points.

On its face, the effect of anti-democratic shocks—shown in Figures 5 and 6—is s more ambiguous. The linear trend in terms of democratic trade loss per year is not significant unless one outlier (Malawi 1969) is removed (Malawi experienced an unexplained negative shock to democratic trade at t-3). But the pattern here is running against the current, since the default is an expected increase per year of 0.43 percent. With a resetting of the baseline to be a 0.43 gain per year rather than zero (and controlling for country effects) pushes the *t*-statistic to |3.28| and the .001 level of significance. With the 0.43 adjustment, the mean shift per year is -0.89 per year during the interval of an anti-democratic slide. The comparable adjustment for the 28 democratic gains is an increase of +1.29 per year.

Another way of analyzing the pro- and anti-democratic shocks is to perform a simple difference-in-differences test. First, we subtract the nation's mean democratic trade over the five pre-shock years from the nation's mean democratic trade over the five post-shock years. This yields 38 scores representing the before-after shock change in democratic trade. Then we compare these difference scores for pro- and anti-democratic shocks. The independent variable predicting mean change in democratic trade is the binary plus or minus polarity of the shock.

The relevant data are shown visually in Figure 7. The nations with pro-democratic shocks gained on average 15.54 percent more democratic trade relative to nations with anti-democratic shocks. This 15.54 coefficient has a *t*-statistic of 3.33 and is significant at the .003 level. Meanwhile, the mean difference in the change in democracy scores for the nations with pro- and anti-democratic shocks is 24 points (out of a possible 42). So, as a rough calculation dividing 15.54 by 24, each unit shift of democracy produces an eventual shift of about two thirds of a percentage point in democratic trade.

This evident effect does not seem large by any standard, but even with the puny N by standards of the trade, this effect is statistically significant. More importantly, this finding appears to be robust to rival interpretations or hidden statistical booby-traps. If nations trade more with democracies when they become democracies than they do when they exit from democratic status, the obvious interpretation is the presence of a causal connection from democracy to trading behavior. Our example illustrates the charm and the simplicity of difference-in-difference analysis.[17]

We leave this section by mentioning an obvious irony. When the degree of confidence in the findings is properly assessed, size of sample does not matter and in fact can be inversely related to the precision of the findings. We have examined dyad-years with

[17] Causal inferences from difference-in-difference analysis of course are not immune from rival interpretations. For instance, sometimes the treatment is induced to correct for low values on the dependent variable, which reequilibrate on their own (regression to the mean). In our example, this would be as if nations became more democratic to induce a more favorable democracy to nondemocracy ratio regarding their trade portfolio. While this scenario is not totally implausible, a large effect of sufficient size to account for the evident democratic trade effect would seem very unlikely. Yet even the mention of this rival scenario illustrates how it is possible to be fooled even by non-experimental studies that seem iron-tight.

over 90,000 cases and nation-years with over six thousand cases and find that properly interpreted, a "significant" democratic trade effect remains elusive. But with 38 cases of actual changes in democratic status, a basic difference-in-difference test shows rather clear evidence of a democratic trade effect.

X. A cross-sectional effect?

Our accounting for the original difference between GKY and their critics was that we suspect that nations might not change their trade patterns much in the short run (consistent with GKY) but still have different long-run dispositions (their fixed effects) that govern their propensity to trade with other democracies. We are now in a position to offer the first (rough) test of this proposition.

The residue from our nation-year analysis contains over 100 dummy variable coefficients representing fixed effects for each nation—the effect of being, for instance, Portugal, on the proportion of its trade with democracies. These estimated nation effects can be influenced by national culture among other factors, and one aspect of culture is the nation's disposition toward democracy. If so, we can imagine that long-term tendency toward democracy conditions a nation to trade with other democracies and attracts other nations to trade with it—all independent of short-term variation in democratic status or even distributive concerns within the polity. Moreover, recent evidence from survey data suggests that non-material interests may affect individuals' disposition towards integration in general and trade in particular.[18]

As a crude test of this cross-sectional "long term" relationship between democracy and democratic trade, we start with the fixed nation and year equation from Table 3. We take the nation dummies for trade (normalized for the arbitrary base year of 1994) and regress them on mean national democracy, adjusting only for the year. The bivariate relationship is shown in Figure 8. Clearly we see a relationship whereby the coefficient predicting democratic trade from democracy is 1.10, which is larger than the companion 0.38 time serial estimate of the democracy effect on trade. The suggestion is that the difference between a maximally despotic nation for a half century to a maximally democratic nation for a half century produces a difference of 22 percentage points in their propensity to trade with democracies.

Of course the cross-sectional relationship is subject to all sorts of rival forces that would require a serious multivariate analysis if it were to become a matter of serious inquiry. Certainly a well specified model would include such variables as a measure of the geographic pull of neighboring autocracies or democracies, the nation's present or past attachment to the British empire or other colonial links, common legal traditions and language, to name a few. Thinking about such a project should renew one's attraction to the time series aspect and the nation-year data set where at least there was little concern

[18] See O'Rourke and Sinnot (2001), Mayda and Rodrik (2005), Hainmuller and Hiscox (2006), Pinto and LeFoulon (2007).

about the influences of third variables unmeasured by nation and year.

XII. Conclusions

This paper has explored the question of democratic trade, one of the controversies in the "Dirty Pool" debate initiated with the publication of Green , Kim, and Yoo (2001). Whereas the central methodological issue of that debate was "fixed effect: yes or no?" we go beyond that. While siding with GKY in favor of fixed effects when analyzing dyads (or nations), our focus was on the units of analysis themselves.

Methodologically, our findings are

- Dyads are inappropriate as units for the democratic trade hypothesis and, we strongly suspect, similar problems in the analysis of trade and other relations among nations. Our randomization tests confirm that the standard errors reported for OLS models of dyadic analysis are wildly optimistic. Estimated standard errors from computer output imply relationships being strongly significant statistically, yet random reshuffling of the independent variables shows that these significance levels are trillions of times too large.
- For testing at least the democratic trade hypothesis, we suggest the nation-year as the unit rather than the dyad-year. And of course this means proper testing. When fixed effects of both nation and year are employed and when the disturbances are allowed to follow an AR1 model (autocorrelation), support for the democratic trade hypothesis is no greater than at the fringe of statistical significance—despite having thousands of observations.
- We then turn to instances of major pro-or anti-democratic shocks to political systems as treatments in a quasi-experimental analysis. Difference-in-difference analysis of 38 such events offers strong evidence that in fact shifts to (from) democracy expand (contract) the nation's proportion of its trade that is with partners that are themselves democratic.

One meta-lesson from our project is that it is possible to be lulled into a false acceptance of a research hypothesis by the force of a large N, which in an imperfect non-experimental session is prone to giving false positives. When the model is wrong, citing significance levels based on the reported standard errors in computer output may only offer the illusion of success.

A second meta-lesson is the need to be humble about findings when different tests give different results. This lesson is offered not only for our readers but also for ourselves.. We have reported a series of results which vary to a dizzying degree regarding how confident we should be in the positive effect of democracy on trade with democracies. Our analysis concludes with a simple model applied to a small number of cases involving democratic or non-democratic interventions. Contrary to some of the previous modeling of this issue, the conclusion from our final analysis is that there is convincing evidence for the democratic trade hypothesis when individual cases experiencing large institutional shocks are put under the microscope. We do not know why we get this result whereas other (presumably inferior) tests involving more observations do not. But as always, it is possible that our confidence is misplaced and we are very wrong.

References

Alt, James and Michael Gilligan (1994). The Political Economy of Trading States: Factor Specificity, Collective Action Problems and Domestic Political Institutions. *Journal of Political Philosophy* 2 (2): 165 – 192.

Beck, Nathaniel, and Johnathan Katz (2001). Throwing Out the Baby with the Bath Water: A Comment on Green, Kim, and Yoon. *International Organization* 55 (2): 487–495.

Bliss, Harry, and Bruce Russett (1998). Democratic Trading Partners: The Liberal Connection, 1962–1989. *The Journal of Politics* 60 (4): 1126 – 47.

Bremer, Stuart (1993). Democracy and militarized interstate conflict, 1816–1965. *International Interactions* 18 (1):231–249.

Donohue, John J. and Justin Wolfers. 2006. "Uses and Abuses of Empirical Evidence in the Death Penalty Debate." *Stanford Law Review* 58: 791-835.

Doyle, Michael (1986). Liberalism and World Politics. *American Political Science Review* Vol. 80 (4): 1151 – 69.

Edgington, Eugene S. and Patrick Onghena. 2007. *Randomization Tests*. Boca Raton, Taylor and Francis Group. 4th ed.

Farber, Henry, and Joanne Gowa (1997). Common interests or common polities: Reinterpreting the democratic peace. *Journal of Politics* 59 (2): 393–417.

Fisher, R.A. 1935. *The Design of Experiments*. Edinburgh: Oliver and Boyd.

Garrett, Geoffrey (2000). The causes of globalization. *Comparative Political Studies* 33 (6–7):941–91.

Green, Donald P., Soo Yeon Kim, and David H. Yoon (2001). Dirty Pool. *International Organization* 55 (2): 441 – 468.

Hainmueller, Jens and Michael J. Hiscox (2006). Learning to Love Globalization: The Effects of Education on Individual Attitudes Toward International Trade. *International Organization* 60 (2): pp 469-498.

Helland, Eric and Alexander Tabarrok. 2004. "Using Placebo Laws to Test 'More Guns, Less Crime'." *Advances in Economic Analysis and Policy* 4: 1-7.

Hiscox, Michael J. (2002). *International Trade and Political Conflict: Commerce, Coalitions, and Mobility*. Princeton, NJ: Princeton University Press.

Jones, Ronald W. (1971). A Three – Factor Model in Theory, Trade and History. In Jagdish Bhagwati, Ronald Jones, Robert Mundell and Jaroslav Vanek, (eds.), *Trade, Balance of Payments and Growth*. Amsterdam: North – Holland.

Manly, Bryan F. J. 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. London, Chapman Hall. 2nd ed.

Mansfield, Edward D., and Rachel Bronson (1997). Alliances, Preferential Trading Arrangements, and International Trade. *American Political Science Review* 91 (1): 94 – 107.

Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff (2000). Free to Trade: Democracies, Autocracies, and International Trade. *American Political Science Review* 94 (2): 305 – 321

Maoz, Zeev, and Bruce Russett (1993). Normative and structural causes of democratic peace. American Political Science Review 87 (3): 624–638.

Marshall, Monty G., Keith Jaggers, and Ted R. Gurr (2008). *Polity IV Project. Political Regime Characteristics and Transitions, 1800 –2008*. Electronic Resource. University of Maryland. URL: http://www.systemicpeace.org/polity/polity4.htm

Mayda, Anna Maria and Dani Rodrik (2005), Why Are Some People (and Countries) More Protectionist Than Others? *European Economic Review*, vol. 49(6): 1393-1430.

Mayer, Wolfgang (1984). Endogenous Tariff Formation. *American Economic Review* 74 (5): 970 – 985.

Moore, David S., Goerge P. McCabe, William M. Duckworth, and Stanley L. Sclove. 2003. *The Practice of Business Statistics Companion Chapter 18: Bootstrap Methods and Permutation Tests*. W.H. Freeman.

Morrow, James D., Randolph M. Siverson, and Tressa E. Tabares (1999). The Political Determinants of International Trade: the Major Powers, 1907 – 1990. *American Political Science Review*, 92 (3): 649 – 661.

Morrow, James D., Randolph M. Siverson, and Tressa E. Tabares (1998). The Political Determinants of International Trade: The Major Powers, 1907–90. *American Political Science Review* 92 (3): 649 – 661.

O'Rourke, Kevin, and R. Sinnott (2001). The Determinants of Individual Trade Policy Preferences: International Survey Evidence. *Brookings Trade Policy Forum 2001*: 157-206.

O'Rourke, Kevin H., and Alan M. Taylor (2006). Democracy and Protectionism. *NBER Working Paper Series No. 12250*, National Bureau of Economic Research.

Oneal, John R. and Bruce Russett (2001). Clear and Clean: The Fixed Effects of the

Liberal Peace. *International Organization* 55 (2): 469 – 485

Oneal, John R., and Bruce Russett (1997). The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950 –1985. *International Studies Quarterly* 41 (2):267 – 93.

Oneal, John R., Frances H. Oneal, Zeev Maoz, and Bruce Russett (1996). The Liberal Peace: I nterdependence, Democracy, and International Conflict, 1950 – 85. *Journal of Peace Research* 33 (1):11–28.

Pinto, Pablo M. and Carmen M. Le Foulon (2007). The Individual Sources of Economic Nationalism: Evidence from Survey Data. *Saltzman Institute of War and Peace Studies Working Paper No. 3* (December 2007).

Rogowski, Ronald (1989). *Commerce and Coalitions: How Trade Affects Domestic Political Alignments*. Princeton, NJ: Princeton University Press.

Samuelson, Paul A. (1971). Ohlin Was Right. *The Swedish Journal of Economics*, 73 (4): 365 – 384.

Stolper, Wolfgang F. and Paul A. Samuelson (1941). Protection and Real Wages. *The Review of Economic Studies* 9 (1): 58 – 73.

Table 1.1: Regression Analysis of Bilateral Trade; Greene, Kim and Yoon Sample, Nominal Dollars

	Pooled	Dyad fixed effects	Pooled with dynamics	Dyad fixed effects with dynamics	Dyad and year fixed effects	Dyad and year fixed effects with dynamics
GDP	1.182 (0.008) *p=.000*	0.810 (0.015) *p=.000*	0.250 (0.006) *p=.000*	0.342 (0.013) *p=.000*	1.688 (0.042) *p=.000*	0.767 (0.036) *p=.000*
Population	-0.386 (0.010) *p=.000*	0.752 (0.082) *p=.000*	-0.059 (0.006) *p=.000*	0.143 (0.068) *p=.035*	1.281 (0.083) *p=.000*	0.317 (0.070) *p=.000*
Distance	-1.342 (0.018) *p=.000*		-0.328 (0.012) *p=.000*			
Alliance	-0.745 (0.042) *p=.000*	0.777 (0.136) *p=.000*	-0.247 (0.027) *p=.000*	0.419 (0.121) *p=.001*	0.459 (0.134) *p=.001*	0.310 (0.121) *p=.010*
Democracy	**0.075** **(0.002)** ***p=.000***	**-0.039** **(0.003)** ***p=.000***	**0.022** **(0.001)** ***p=.000***	**-0.009** **(0.002)** ***p=.000***	**-0.015** **(0.003)** ***p=.000***	**-0.0003** **(0.003)** ***p=.902***
Lagged bilateral trade			0.736 (0.002) *p=.000*	0.533 (0.003) *p=.000*		0.521 (0.003) *p=.000*
Constant	-17.331 (0.265) *p=.000*	-47.994 (1.999) *p=.000*	-3.046 (0.177) *p=.000*	-13.745 (1.676) *p=.000*	-102.229 (2.706) *p=.000*	-40.706 (2.461) *p=.000*
	N=93,924	N=3,079 T>=20	N=88,946	N=3,079 T>=20	N=3,079 T>=20	N=3,079 T>=20
R^2	0.36	0.63	0.73	0.76	0.65	0.62

Democracy is the lower value within the dyad.
GDP, population, distance, and bilateral trade are natural log transformed.

Table 1.2: Regression Analysis of Bilateral Trade; Greene, Kim and Yoon Sample, Real 1996 Dollars

	Pooled	Dyad fixed effects	Pooled with dynamics	Dyad fixed effects with dynamics	Dyad and year fixed effects	Dyad and year fixed effects with dynamics
GDP	0.966	0.544	0.179	0.212	1.688	0.767
	(0.008)	(0.016)	(0.005)	(0.013)	(0.042)	(0.036)
	p=.000	*p=.000*	*p=.000*	*p=.000*	*p=.000*	*p=.000*
Population	-0.194	0.778	-0.003	0.161	1.281	0.317
	(0.010)	(0.082)	(0.006)	(0.068)	(0.083)	(0.070)
	p=.000	*p=.000*	*p=.607*	*p=.018*	*p=.000*	*p=.000*
Distance	-1.346		-0.314			
	(0.018)		(0.012)			
	p=.000		*p=.000*			
Alliance	-0.832	0.808	-0.260	0.424	0.459	0.310
	(0.043)	(0.136)	(0.027)	(0.121)	(0.134)	(0.121)
	p=.000	*p=.000*	*p=.000*	*p=.000*	*p=.001*	*p=.010*
Democracy	**0.095**	**-0.042**	**0.027**	**-0.009**	**-0.015**	**-0.0003**
	(0.002)	**(0.003)**	**(0.001)**	**(0.002)**	**(0.003)**	**(0.003)**
	p=.000	***p=.000***	***p=.000***	***p=.000***	***p=.000***	***p=.902***
Lagged bilateral trade			0.748	0.536		0.521
			(0.002)	(0.003)		(0.003)
			p=.000	*p=.000*		*p=.000*
Constant	-12.311	-35.383	-1.563	-7.818	-100.563	-40.679
	(0.271)	(2.005)	(0.176)	(1.675)	(2.706)	(2.461)
	p=.000	*p=.000*	*p=.000*	*p=.000*	*p=.000*	*p=.000*
	N=93,924	N=3,079 T>=20	N=88,946	N=3,079 T>=20	N=3,079 T>=20	N=3,079 T>=20
R^2	0.31	0.62	0.72	0.69	0.66	0.60

Democracy is the lower value within the dyad.
GDP and bilateral trade are in real 1996 dollars and are natural log transformed.
Population and distance are natural log transformed.

Table 2: Regression Analysis of Bilateral Trade, 1950-2000

	Pooled	Dyad fixed effects	Pooled with dynamics	Dyad fixed effects with dynamics	Dyad and year fixed effects	Year fixed effects with dynamics	Dyad and year fixed effects with dynamics
GDP	0.911	0.896	0.077	0.199	0.918	0.087	0.224
	(0.003)	(0.005)	(0.002)	(0.004)	(0.007)	(0.002)	(0.005)
	p=.000	*p=.000*	*p=.000*	*p=.000*	*p=.000*	*p=.000*	*p=.000*
Population	0.592	0.190	0.052	0.040	0.252	0.057	0.103
	(0.002)	(0.007)	(0.001)	(0.004)	(0.012)	(0.001)	(0.007)
	p=.000	*p=.000*	*p=.000*	*p=.000*	*p=.000*	*p=.000*	*p=.006*
Distance	-0.751		-0.063			-0.065	
	(0.006)		(0.002)			(0.002)	
	p=.000		*p=.000*			*p=.000*	
Alliance	0.351	0.378	0.028	0.065	0.384	0.024	0.064
	(0.014)	(0.024)	(0.005)	(0.015)	(0.024)	(0.005)	(0.015)
	p=.000	*p=.000*	*p=4e-7*	*p=9e-6*	*p=.000*	*p=1e-5*	*p=1e-5*
Democracy	**0.039**	**0.016**	**0.003**	**0.003**	**0.013**	**0.004**	**0.003**
	(0.001)	**(0.001)**	**(0.000)**	**(0.000)**	**(0.001)**	**(0.000)**	**(0.000)**
	p=.000	***p=.000***	***p=.000***	***p=.000***	***p=.000***	***p=.000***	***p=2e-12***
	rand p=.001	***rand p=0.011***	***rand p=.014***	***rand p=.013***	***rand p=.036***	***rand p=.001***	***rand p=0.022***
Lagged bilateral trade			0.923	0.791		0.920	0.791
			(0.001)	(0.002)		(0.001)	(0.002)
			p=.000	*p=.000*		*p=.000*	*p=.000*
Constant	-17.580	-15.953	-1.498	-3.509	-17.312	-1.655	-5.161
	(0.090)	(0.097)	(0.039)	(0.067)	(0.259)	(0.043)	(0.178)
	p=.000	*p=.000*	*p=.000*	*p=.000*	*p=.000*	*p=.000*	*p=.000*
	N=119,640	N=2,346	N=117,294	N=2,346	N=2,346	N=2,346	N=2,346
		T=51		T=50	T=51	T=50	T=50
R^2	0.66	0.87	0.95	0.95	0.87	0.95	0.95

Democracy is the lower value within the dyad.
GDP and bilateral trade are in real 1996 dollars and are natural log transformed.
Population and distance are natural log transformed.
A *p*-value equal to zero indicates a value smaller than 2.2e-16.

Table 3: Modeling Democratic Trade as a Function of a Nation's Degree of Democracy

Dependent Variable = Percent of Trade with Democracies	*All Cases, NxT=6297*			*Balanced Panel Only 51-Year Time Series NxT=3519*		
	b	*st.err*	*p-value*	*b*	*st.err*	*p-value*
Fixed N,T	0.38	0.04	.000	0.63	0.05	.000
Fixed N	0.88	0.04	.000	1.09	0.05	.000
Random	0.90	0.04	.000	1.10	0.05	.000
Fixed N,T, AR1	0.12	0.15	.017	0.07	0.05	.128
Fixed N,T, with Dynamics	0.08	0.02	.001	0.08	0.02	.000
Pooled	1.15	0.03	.000	1.34	0.04	.000
rho (AR1)	0.87			0.92		
Lag Coefficient (Dynamic)	0.84			0.92		

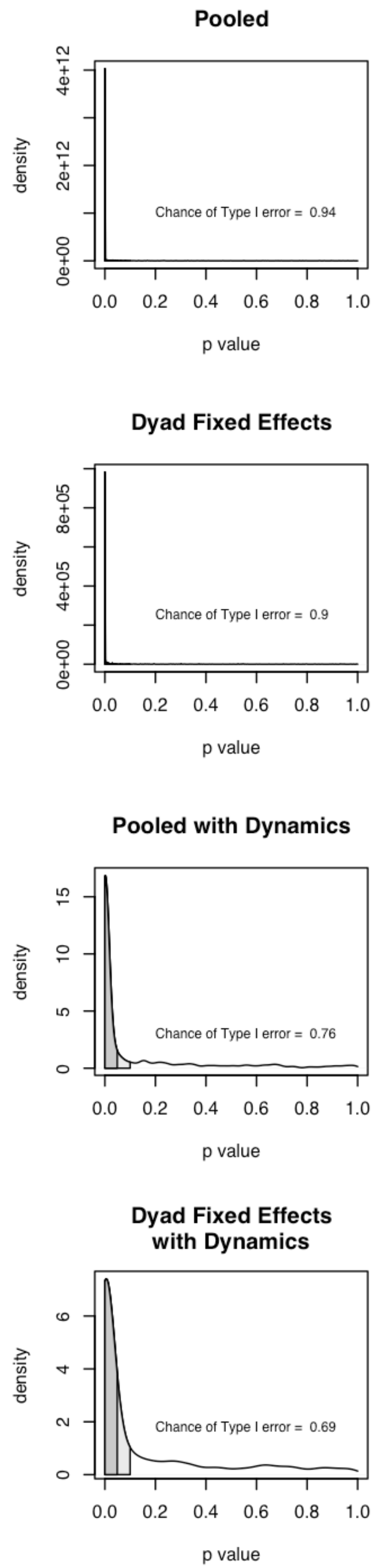

Figure 1.1: Randomization Test *p*-values

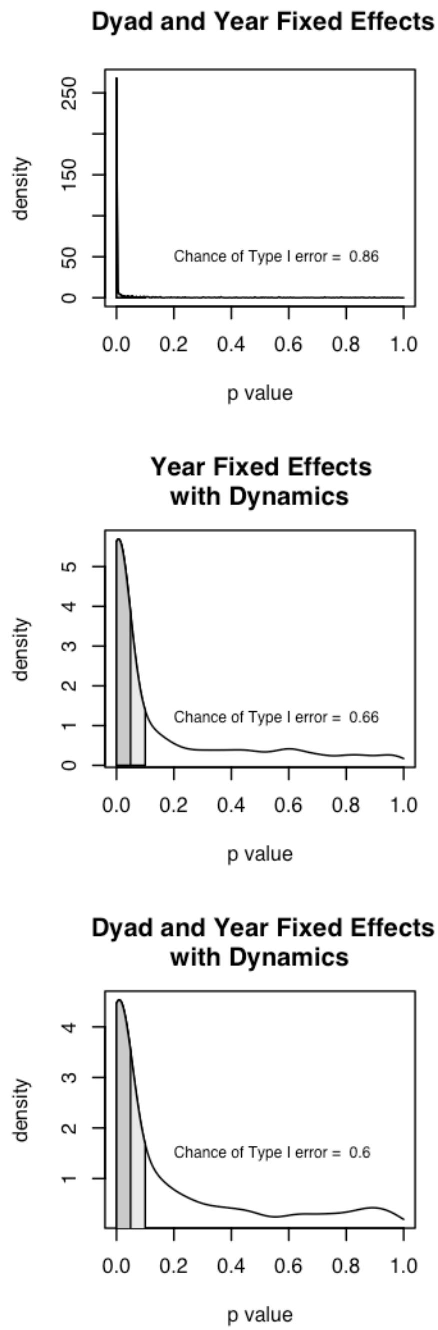

Figure 1.1: Randomization Test *p*-values (continued)

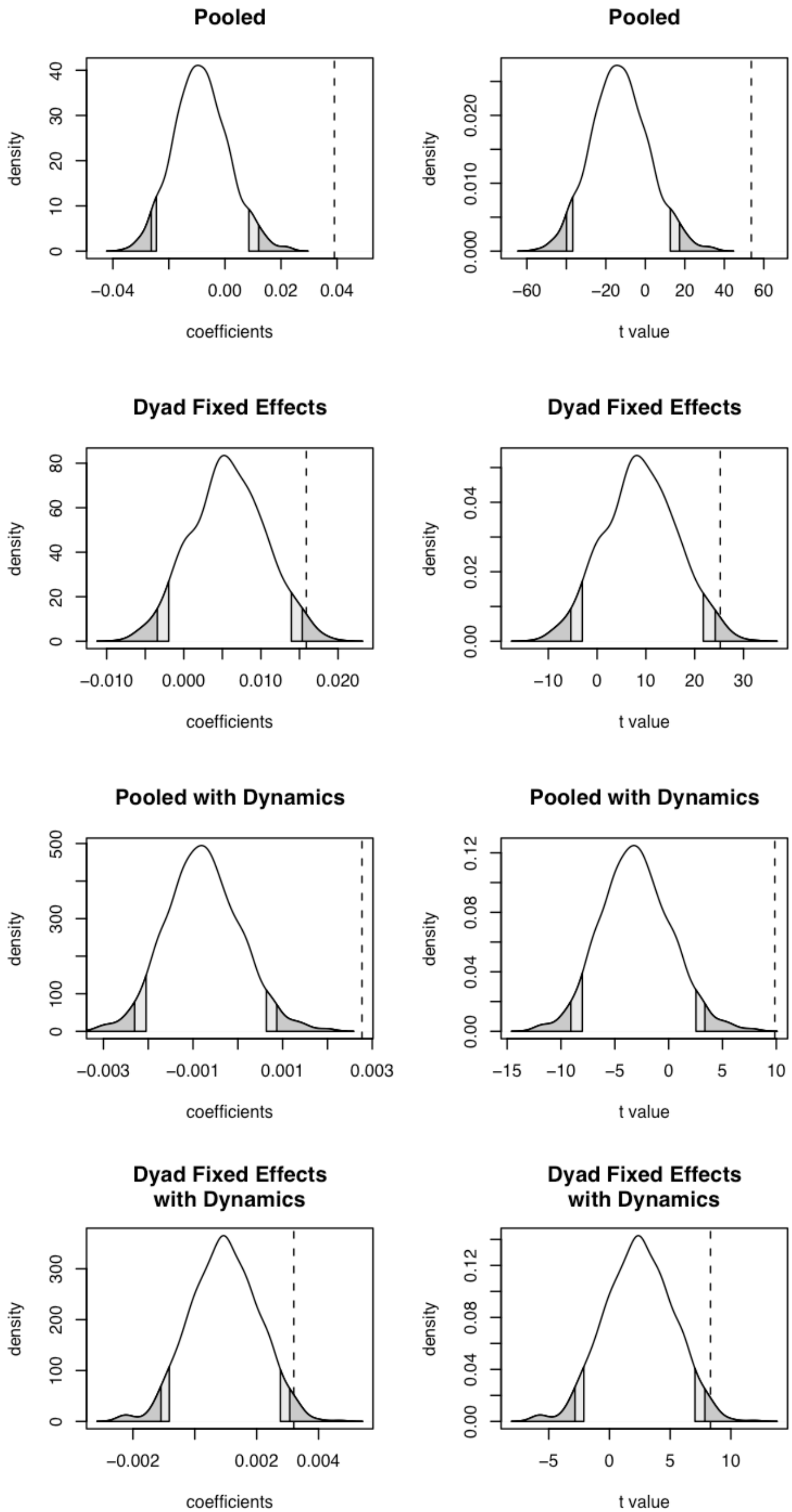

Figure 1.2: Randomization Test Results

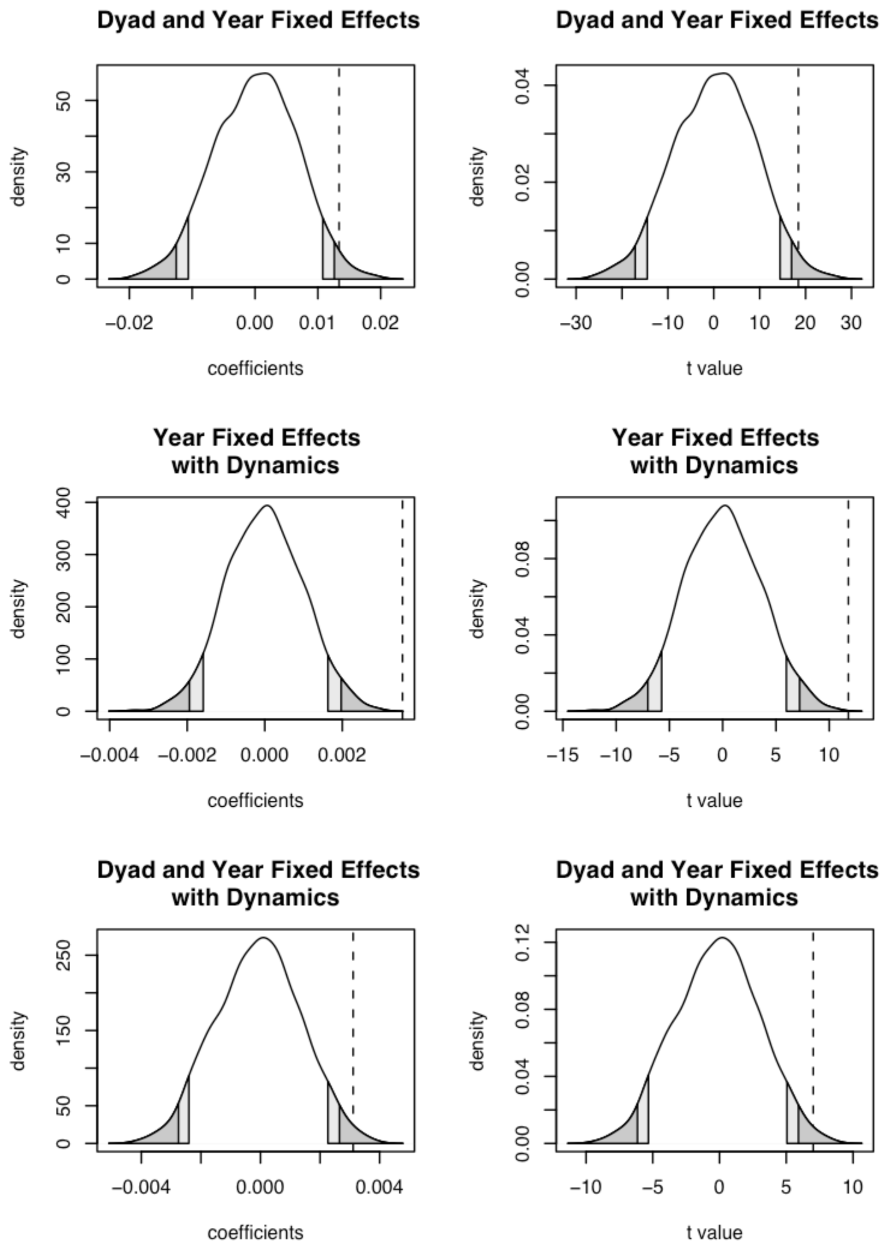

Figure 1.2: Randomization Test Results (continued)

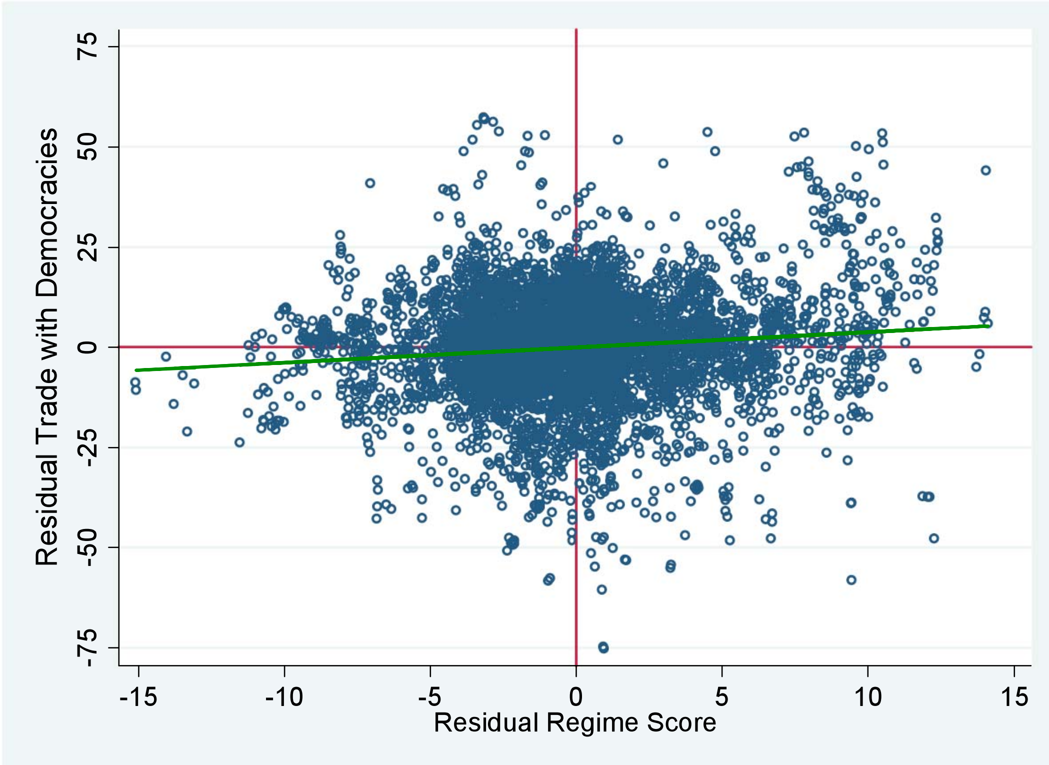

Figure 2. Residual proportion of trade with other democracies by residual regime (democracy) score.
Data are nation-years. Controls are for nation and year dummies.

$Y = .03 + .38X + e$

$st.err. = .044$

$p = .000$

$N = 6459$

$Adj.Rsq = 0.011$

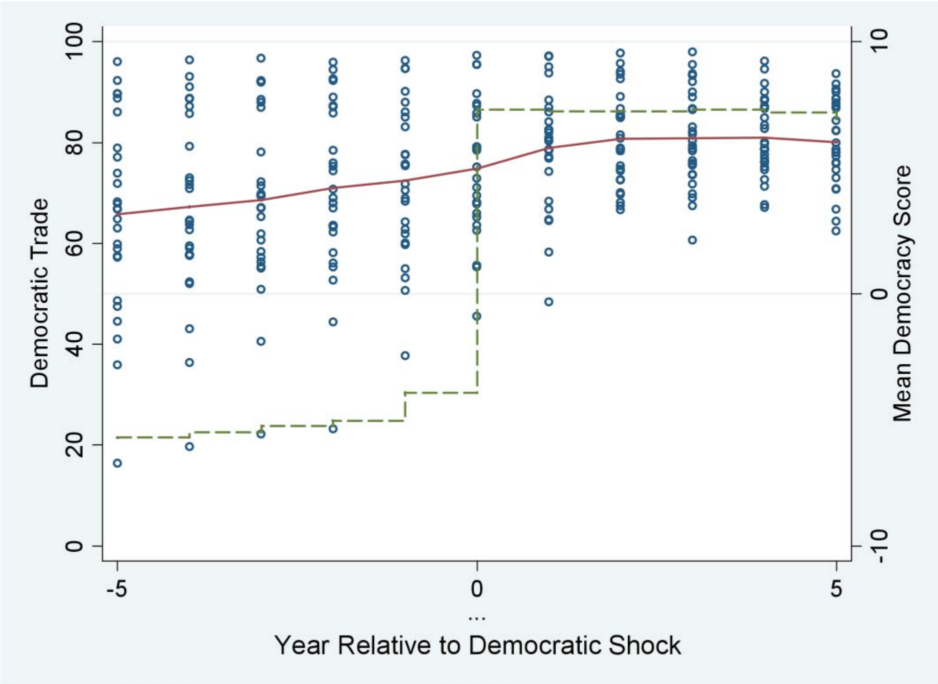

Figure 3. Democratic trade (left scale) by year relative to the democratic shock at year 0.
Mean democracy scores (right scale) are represented by the dashed line.

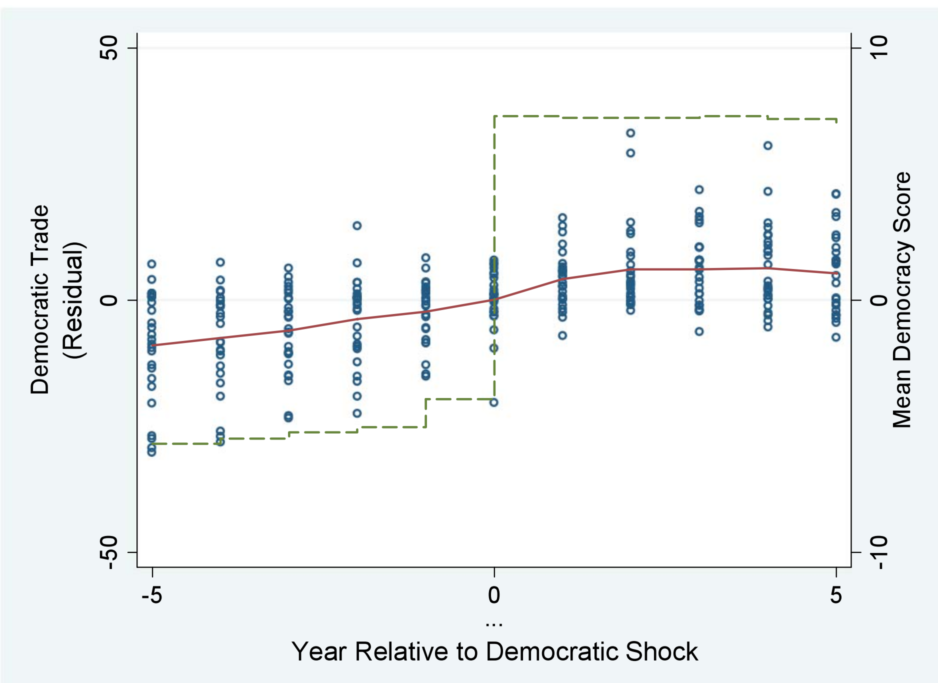

Figure 4. Residualized democratic trade (left scale) by year relative to the democratic shock at year 0.
Mean democracy scores (right scale) are represented by the dashed line.

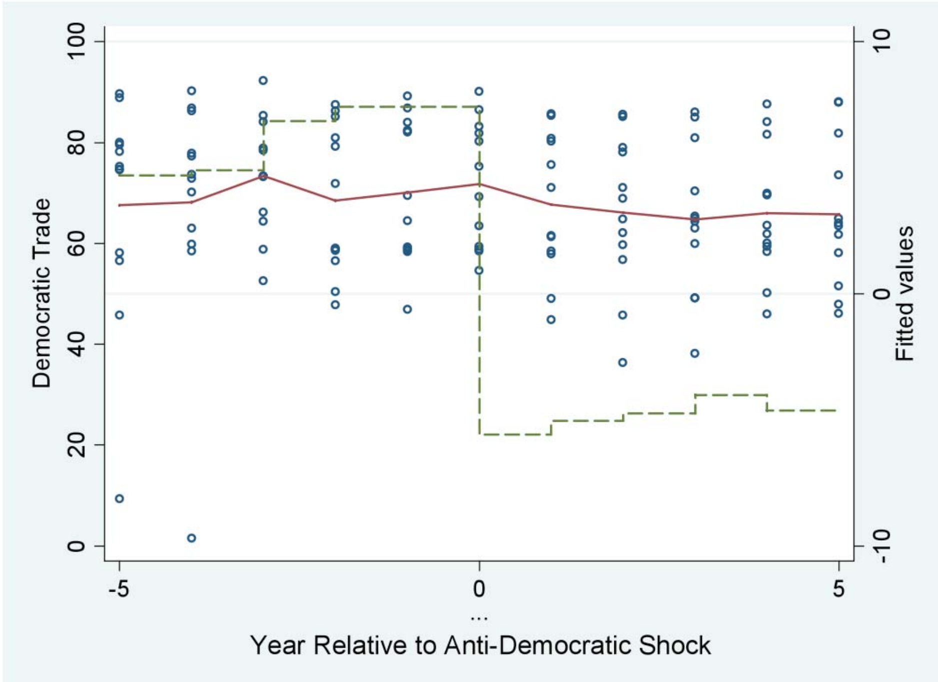

Figure 5. Democratic trade (left scale) by year relative to the anti-democratic shock at year 0.
Mean democracy scores (right scale) are represented by the dashed line.

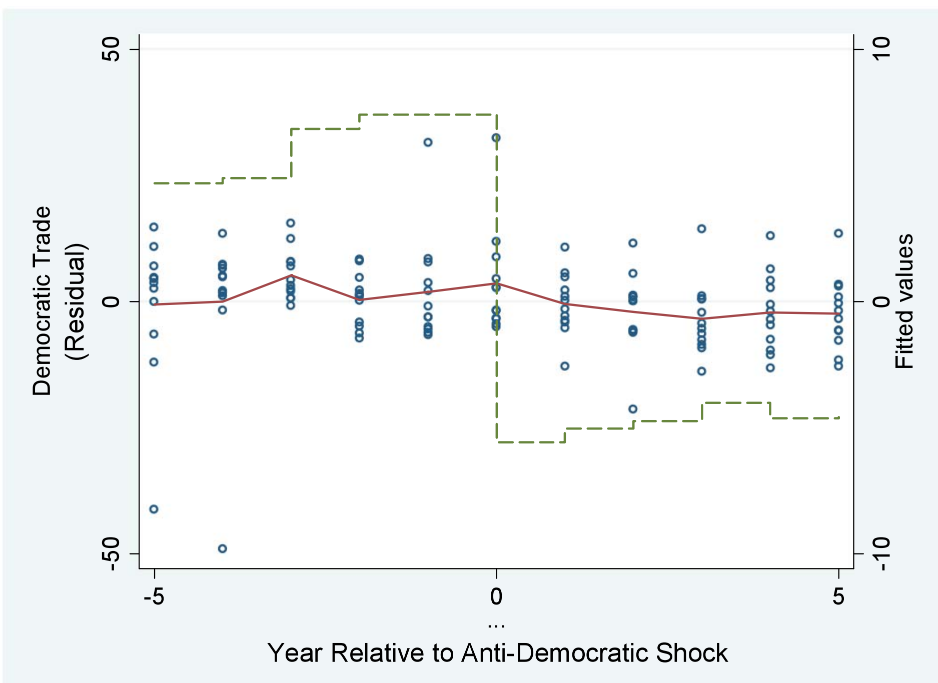

Figure 6. Residualized democratic trade (left scale) by year relative to the anti-democratic shock at year =0.
Mean democracy scores (right scale) are represented by the dashed line.

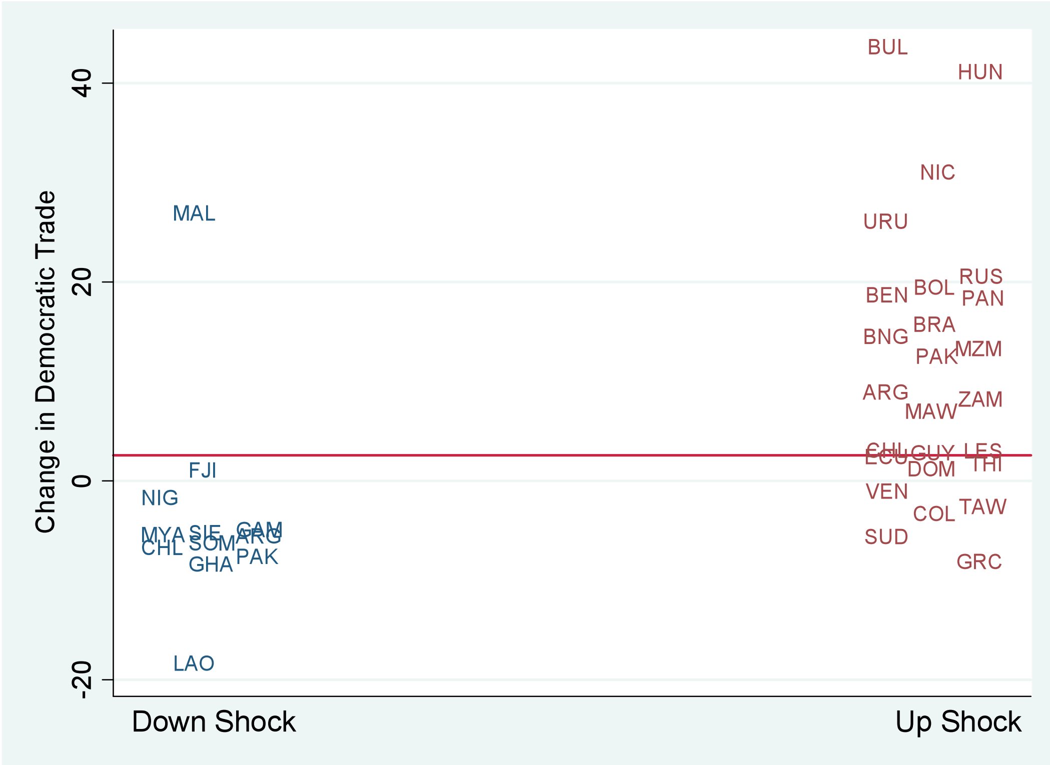

Figure 7. Change in Democratic Trade as a function of Democratic Shocks. Change is mean democratic trade over the five post-shock years minus democratic trade over the five pre-shock five years. The horizontal line at 2.6 represents the mean tendency toward democratic trade per year x 6 years, the time between the midpoint of the pre-event sample and the midpoint of the post-event sample

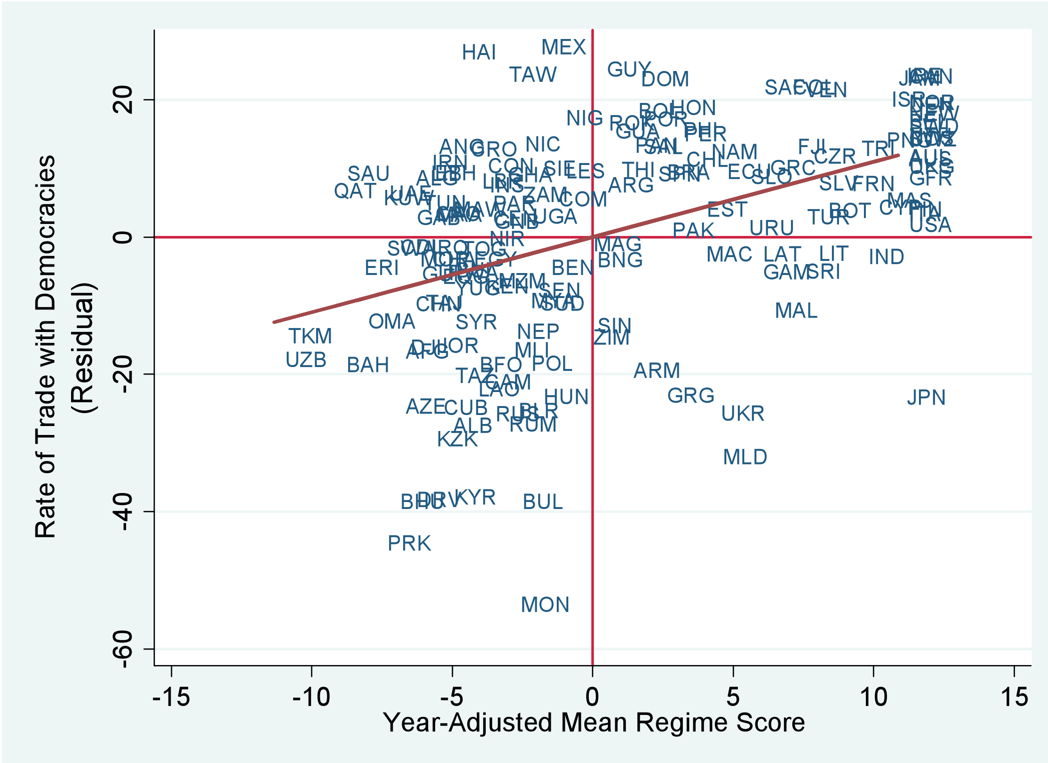

Figure 8. Residual (macro level) democratic trade by mean regime (democracy) scores adjusted for years in data set.
Data are nations.

$Y = 1.10X + e$

$st.err. = .185$

$p = .000$

$N = 154$

$Adj.Rsq = 0.183$