DYADIC ANALYSIS IN INTERNATIONAL RELATIONS: A CAUTIONARY TALE

Robert S. Erikson
Department of Political Science
ColumbiaUniversity
rse14@columbia.edu

Pablo M. Pinto
Department of Political Science
Columbia University
pp2162@columbia.edu

Kelly T. Rader
Department of Political Science
Yale University
kelly.rader@yale.edu

November 25, 2013

Supplementary Appendices

Appendix I - Data description and sources

Trade: is the natural log of bilateral trade in constant 1996 dollars. Trade is measured in million dollars, hence we add 1 (equivalent to 1 million dollars) to the reported trade values before taking the natural log, to avoid losing observations with no trade. The source for the trade data is Gledistch (2002).
http://privatewww.essex.ac.uk/~ksg/exptradegdp.html

GDP: is the sum of the natural log of real GDP per capita of countries A and B in constant 1996 dollars. The source for this data is Gledistch (2002).

Population: is the sum of the natural log of the population of countries A and B. The source for this data is Gledistch (2002).

Distance: is the natural log of the distance in miles between capitals or the shortest distance between ports for large countries (US, USSR/Russia, Canada). The source for this data is Bennet and Stam (2000), Eugene Dataset, available at:
http://eugenesoftware.org.

Allied: is a dummy variable which takes a value of 1 if the dyad members are directly and formally allied, and zero otherwise. The source for this data is Gibler and Sarkees (2004) and Gibler (2008) Correlates of War Alliance v.3.03, available at:
http://www.correlatesofwar.org/COW2 Data/Alliances/alliance.htm

Minimum Democracy: is the lowest regime score for the dyad members. The source for this data is Marshall et al (2008) Polity IV, available at http://www.systemicpeace.org.

Works Cited

Bennett, D. Scott, and Allan Stam (2000). "EUGene: A Conceptual Manual." *International Interactions* 26:179-204. Website: http://eugenesoftware.org

Gibler, Douglas M., and Meredith Sarkees (2004). "Measuring Alliances: The Correlates of War Formal Interstate Alliance Data set, 1816-2000." *Journal of Peace Research* 41(2): 211-222.

Gibler, Douglas M.. (2008). *International Military Alliances from 1648 to 2008*. Washington: Congressional Quarterly Press.

Gleditsch, Kristian S. (2002). "Expanded Trade and GDP Data." *Journal of Conflict Resolution* 46: 712-24.

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

Appendix II: Democracy and Trade

One long-standing question in international relations is whether democracies trade more with other democracies. IR scholars have focused on the role of democratic governance on trade, in a development that parallels the analysis of the link between democracy and international conflict and cooperation.[1]

The democratic trade hypothesis is derived from the intuition that the quality and similarity of domestic institutions are associated with lower barriers to international transactions. Democratic governance is associated with the rule of law and credibility of commitments that help sustain stable international commerce regimes (Schumpeter 1950, Doyle 1983, 1986, Oneal and Russett 1997, 2001b). Moreover, democratic governments are thought to be more reliable trading partners (Polachek 1997, Mansfield, Milner and Rosendorff 2000, Przeworski et al. 2000, Polachek and Seigle 2007). Additionally, consumers are willing to pay a higher premium for goods and services originating in countries with similar cultural traits and political affinity (Noland 2005).

Following the lead of the literature on the democratic peace, empirical analyses of the democratic trade hypothesis rely on dyadic data, which present several statistical challenges. Based on a search of nine prominent journals in Political Science and International Relations over the past two decades, we find 192 articles that relied on dyadic analysis. Among these 192 articles, 29 were on issues related to international

[1] See Doyle (1983, 1986) for the philosophical roots of the democratic peace. Bremer (1992, 1993), Maoz and Russett (1993), Oneal, Oneal, Maoz, and Russett (1996), Oneal and Russett (1997) are prominent examples of empirical tests using dyads as units of analysis.

trade, where democratic governance enters the right hand side of the estimating equation as an explanatory or control variable.[2]

In Table A.II.1 we present the methods used in those 29 analyses, number of observations and *t*-statistics for the relevant measure of democracy. The practice is also found in economic journals: Rose (2004, 2007) and Tomz, Goldstein, and Rivers (2007), for instance, also use dyadic data in their analyses of the effect of GATT membership on trade.

[2] Among empirical studies of the link between democracy and trade the following use dyadic datasets with 100,000 or more observations: Mansfield, Milner, and Rosendorff (2002), Mansfield and Reinhardt (2003, 2008), Bayer and Rupert (2004), Gowa and Kim (2005), Blanton and Blanton (2007), Goldstein, Rivers, and Tomz (2007), Kono (2008), Long (2008), Bartilow and Voss (2009), Johnson, Souva, and Smith (2013).

Table A.II.1 - Recent quantitative analyses of bilateral trade appearing in nine leading journals (1993-2013)[3]

II-1. Analyses with democracy as main independent variable				
Author/Year	DV (years covered)	IV/*t-stat*	N	Models
Bliss and Russett (1998)	Log of dyadic trade volume (1962-1989)	- Democracy, the weak-link measurement [0.39,4.3]	882 for each year	- Undirected dyad - Separate OLS estimation for each year - Huber-White standard error
Morrow et al (1998)	Exports from state *i* to state *j* (1907-1990)	- The presence of militarized dispute - Common interests - Joint democracy dummy [9.3] - Alliances	2,631	- Directed dyad-year - OLS estimation with panel-corrected standard errors - Prais-Winston treatment
Morrow et al (1999)	Exports from state *i* to state *j* (1907-1990)	- The presence of militarized dispute - Common interests - Joint democracy dummy [2.9] - Alliances	2,631	- Directed dyad-year - Prais-Winston estimation with panel-corrected standard errors
Mansfield et al (2000)	Log of the value of exports from state *i* to state *j* (1960-1990)	- Mixed regime pair [-6.0, -4.4] - Autocratic regime pair [-1.4, 0.5] - Other incoherent regime pair [-3.3,-2.3]	30,480-33,116	- Directed dyad-year - Pooled OLE estimation - White heteroskedasticity-consistent standard error - Country FE, year FE.
Mansfield et al (2002)	Formation of PTA (1951-1992)	- Regime type, Polity III - Regime type of *i* [Z-Stat: 8.8,8.9] - Regime type of *j* [Z-Stat: 8.2,8.5]	223,568	- Directed dyad-year - Logistic regression with a natural spline function included
Kono (2008)	Log of country *i* 's imports from country *j* as a percentage of *i*'s GDP (1950-2000)	- Democracy *i* - Democracy *j* - Democracy *i* X ratio *ij*	543,331-780,754	- Dyad-year - Error correction model - Robust-cluster estimators clustered at dyadic level *Robustness:* country FE
Bartilow and Voss (2009)	Log of the value of exports from state *i* to state *j* (1948-1997)	- Joint democracy [-0.8] - Exporter's level of democracy [16.2] - Importer's level of democracy [16.6]	255,600	- Directed dyad-year - OLS estimation with panel corrected standard errors - Prais-Winsten transformation *Robustness*: year FE

[3] List of journals: American Journal of Political Science; American Political Science Review; International Interactions; International Organization; International Studies Quarterly; Journal of Conflict Resolution; Journal of Peace Research; Journal of Politics; World Politics.

II-2. Analyses with democracy as a control variable				
Author/Year	DV	IV/T-stat	N	Models
Li and Sacko (2002)	Log of dyadic trade volume (1950-1992, 1870-1992)	- Dyadic dispute - *Control*: Joint democracy dummy [5.6,7]	40,373-81212	- Undirected dyad-year - Two way fixed effects least square estimator with Huber robust standard errros - Dyadic FE, year FE
Long (2003)	Log of dyadic trade volume among major powers (1885-1990)	- Defense-pact alliance, non-defense-pact alliance - *Control*: Joint democracy dummy [2.9, 3.0]	1,281	- Undirected dyad-year - GLS (Prais-Winsten) with panel-corrected standard errors
Bayer and Rupert (2004)	Log of dyadic trade volume (1950-1992)	- Occurrence of civil war - *Control*: Joint democracy dummy [Coefficient: 0.000, Standard errors: 0.000]	101,441	- Undirected dyad-year - OLS estimation with robust standard errors clustered on the country-dyad. *Robustness*: Two-way fixed effects model
Gowa and Mansfield (2004)	Log of exports from state *i* to state *j* (1907-1991, major power)	- Alliances - *Control:* Joint democracy [8.3, 9.3]	3,209-3,774	- Directed dyad-year - Prais-Winsten estimation with panel-corrected standard errors
Keshk et al (2004)	Log of dyadic trade volume	- Dispute - *Control:* Log of lower democracy score [9.3, 10]	143,792	- Undirected dyad-year - OLS estimation with standard errors from i) Maddala procedure and ii) RSE cluster procedure
Simmons (2005)	Log of dyadic trade volume (1950-1995)	- Territorial dispute - *Control*: Joint democracy [1.5, 1.8]	14,362-14,779	- Contiguous dyad-year - OLS estimation with robust standard errors clustered by dyad - Country FE
Long and Leeds (2006)	Log of dyadic trade volume before WWII (1885-1938, 1920-1938, 1885-1913)	- Economically linked alliance, non-linked alliance - *Control*: joint democracy dummy [2.6, 5.6]	3,337-9,342	- Undirected dyad-year - FGLS (Prais-Winsten) regression with panel-corrected standard errors
Blanton and Blanton (2007)	Imports from state *i* to state *j* (1989-2000)	- Human rights indicators - *Control*: lower democracy score [15.0, 23.0]	157,349	- Directed dyad-year - FGLS estimation with Huber-White standard errors clustered at dyad
Kastner (2007)	Log of dyadic trade volume (1960-1992)	Conflicting interests *Control:* Joint democracy dummy [-0.43, 0.48]	57,472-60,809	- Undirected dyad-year - OLS estimation with Huber/White robust standard errors clustered at dyadic level
Long (2008)	Log of exports from state *i* to state *j* (1984-1997)	- Armed conflict, political conflict risk - *Control:* Joint democracy dummy [10.5]	217,340	- Directed dyad-year - FGLS (Prais-Winsten) regression with panel-corrected standard errors

Souva, Smith and Rowan (2008)	Log of dyadic trade volume (1960 - 1999)	- Market promoting institutions, democracy, interaction of market promoting institutions and democracy - Base model without interaction, Democracy [8.5]	48,642-96,844	- Undirected dyad-year - OLS estimation with panel-corrected standard errors *Robustness*: Dyadic FE, GEE
Powell and Rickard (2010)	Log of dyadic trade volume (1955-1998)	- 7 binary variables of legal systems of both states in a dyad - *Control*: Joint democracy dummy [-12.6]	94,821	- Undirected dyad-year - OLS estimation with panel-corrected standard errors - *Robustness*: Dyadic FE
	Log of country's total trade volume (1955-2000)	- 3 binary variables of legal systems *Control*: Democracy [26.7]	4,239	- Country-year - OLS estimation with robust standard errors
Tobin and Busch (2010)	Formation of North-South PTA	- Bilateral investment treaty *Control*: Democracy (Polity score of host country) [Z-stat: 0.2]	97,108	- Undirected dyad-year - Logistic regression with robust standard errors clustered by country - Rare events correction - Natural spline with 5 knots included - Year FE
Manger (2012)	Formation of North-South PTA (1995-2007)	- Vertical intra-industry trade (VIIT share) *Control:* Democracy/North [Z-stat: -4.0,-3.6], democracy/South [2.5,5.29]	11,490-12,623	- Two stage instrumental variable approach - Logistic regression with standard errors bootstrapped by resampling from panels with 1,000 repetitions - Year FE in the first stage - *Robustness*: Country FE, dyadic FE in the first stage
Saikawa (2013)	Log of dyadic automobile exports trade value	- Emission standards *Control*: Joint democracy dummy [0.9]	47,322	- Directed dyad-year - OLS estimation with robust standard errors clustered by dyads - Country FE, year FE
Gowa and Hicks (2013)	Log of imports (1919-1938)	-Trade blocs *Control*: Joint democracy dummy [1.9, 2.0]	35,199	- Directed dyad-year - OLS estimation with standard errors clustered by dyads - Dyad FE, importer/export year dummy

II.3. Analyses with international institutions as main independent variable				
Author/Year	DV	IV/T-stat	N	Models
Mansfield and Bronson (1997)	Log of the real value of exports from state *i* to state *j* (1960-1990)	- Joint PTA dummy [12.6,13.1] - Joint PTA X Alliance [9.1,9.2] - Joint PTA X Major power [-8.4,-7.9] - *Control*: Joint GATT dummy [1.1,1.9]	30,418-32,156	- Directed dyad-year - OLS estimation with White heteroskedasticity-consistent standard errors - Country FE, year FE
Mansfield and Reinhardt (2003)	Formation of PTA (1948-1998)	- Number of members in GATT/WTO [8.4, 12.9] - MTN round underway [10.9, 12.3] - *Control*: democracy [9.4 , 11.8]	149,308-259,267	- Directed dyad-year - Logistic regression with robust standard errors clustered by undirected dyad - Six duration splines included
Rose (2004)	Log of the average value of imports and exports between *i* and *j* (1948-1999)	- Joint GATT/WTO dummy [-3,3] - One GATT/WTO dummy [-3.3,1.3]	114,615-234,597	- Undirected dyad-year - OLS estimation with standard errors clustered by dyad - Year FE *Robustness*: GLS, dyadic FE, Prais-Winsten
Gowa and Kim (2005)	Log of imports from state *i* to state *j* (1950-1994)	- Both GATT members dummy [-0.4,7.7] - One GATT member dummy [-0.6,3.7] - *Control*: Joint democracy [8.4,9.2]	267,970-278,328	- Directed dyad-year - OLS estimation with Newey-West standard errors clustered by dyad - Dyadic FE, year FE
Goldstein, Rivers and Tomz (2007)	Log of imports from state *i* to state *j* (1946-2004)	- Both formal GATT/WTO members dummy [-2.33, 3.5] - One GATT/WTO member dummy [-7, -1]	381,656	- Directed dyad-year - OLS estimation with robust standard error clustered by directed dyad - Dyadic FE, year FE
Mansfield and Reinhardt (2008)	Trade volatility (1951-2001)	- Joint GATT/WTO dummy [-3.8,-3.3] - Joint PTA [-5.8,-2.4]	473,797-527,883	- Directed dyad-year - OLS estimation, conditional logit, heteroskedastic regression, ARCH-in-mean models - Dyadic FE, year FE
Johnson et al (2013)	Log of the average of the total imports and total exports between *i* and *j* (1948-1999)	- Joint GATT/WTO membership [-2.1,0.6] - One GATT/WTO membership [1.5,2.3] - Market promoting institutions (MPI), the weak-link measurement [3.8,5.0] - GATT/WTO X MPI [3.5,4.5]	185,899	- Undirected dyad-year - OLS estimation with robust standard errors clustered on the dyad - Dyadic FE, year FE

Works Cited

Bartilow, Horace A., and D. Stephen Voss. "Market Rules: The Incidental Relationship between Democratic Compatibility and International Commerce." *International Studies Quarterly* 53.1 (2009): 103-124.

Bayer, Reşat, and Matthew C. Rupert. "Effects of civil wars on international trade, 1950-92." *Journal of Peace Research* 41.6 (2004): 699-713.

Blanton, Robert G., and Shannon Lindsey Blanton. "Human Rights and Trade: Beyond the "Spotlight"." *International Interactions* 33.2 (2007): 97-117.

Bliss, Harry, and Bruce Russett. "Democratic trading partners: the liberal connection, 1962–1989." *The Journal of Politics* 60.4 (1998): 1126-1147.

Dixon, William J., and Bruce E. Moon. "Political Similarity and American Foreign Trade Patterns." Political Research Quarterly Vol. 46, no. 1 (1993):5-25.

Doyle, Michael W.. "Kant, Liberal Legacies, and Foreign Affairs." Philosophy and Public Affairs Vol. 12, no. 3 (1983): 205–235

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

Goldstein, Judith, L., Douglas Rivers, and Michael Tomz. "Institutions in International Relations: Understanding the Effects of the GATT and the WTO on World Trade." *International Organization* 61.1 (2007): 37-67.

Gowa, Joanne and Edward D. Mansfield. "Alliances, imperfect markets, and major-power trade." *International Organization* 58.4 (2004): 775-805.

Gowa, Joanne, and Soo Yeon Kim. "An exclusive country club: the effects of the GATT on trade, 1950–94." *World Politics* 57.4 (2005): 453-478.

Gowa, Joanne and Raymond Hicks. "Politics, Institutions and Trade: Lessons of the Interwar Era." *International Organization* 67.3 (2013): 439-467.

Johnson, Jesse C., Mark Souva, and Dale L. Smith. "Market- Protecting Institutions and the World Trade Organization's Ability to Promote Trade." *International Studies Quarterly* (2013): *Online First View*

Kastner, Scott L. "When Do Conflicting Political Relations Affect International Trade?" *Journal of Conflict Resolution* 51.4 (2007): 664-688.

Keshk, Omar MG, Brian M. Pollins, and Rafael Reuveny. "Trade still follows the flag: The primacy of politics in a simultaneous model of interdependence and armed conflict." *Journal of Politics* 66.4 (2004): 1155-1179.

Kono, Daniel Yuichi. "Democracy and trade discrimination." *The Journal of Politics* 70.4 (2008): 942-955.

Li, Quan, and David H. Sacko. "The (Ir)relevance of Militarized Interstate Disputes for International Trade." *International Studies Quarterly* 46.1 (2002): 11-43.

Long, Andrew G. "Defense pacts and international trade." *Journal of Peace Research* 40.5 (2003): 537-552.

Long, Andrew G. "Bilateral trade in the shadow of armed conflict." *International Studies Quarterly* 52.1 (2008): 81-101.

Long, Andrew G., and Brett Ashley Leeds. "Trading for Security: Military Alliances and Economic Agreements." *Journal of Peace Research* 43.4 (2006): 433-451.

Manger, Mark S. "Vertical Trade Specialization and the Formation of North-South PTAs." *World Politics* 64.4 (2012): 622-658.

Mansfield, Edward D., and Rachel Bronson. "Alliances, preferential trading arrangements, and international trade." *American Political Science Review* 91.1 (1997): 94-107.

Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff. "Free to trade: Democracies, autocracies, and international trade." *American Political Science Review* 94.2 (2000): 305-321.

Mansfield, Edward, Helen Milner, and Peter Rosendorff. "Why democracies cooperate more: Electoral control and international trade agreements." *International Organization* 56.3 (2002): 477-513.

Mansfield, Edward D. and Eric Reinhardt. "Multilateral determinants of regionalism: The effects of GATT/WTO on the formation of preferential trading arrangements." *International Organization* 57.4 (2003): 829-862.

Mansfield, Edward D., and Eric Reinhardt. "International institutions and the volatility of international trade." *International Organization* 62.4 (2008): 621-652.

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

Morrow, James D., Randolph M. Siverson, and Tressa E. Tabares. "The political determinants of international trade: The major powers, 1907-90." *American Political Science Review* 92.3 (1998): 649-661.

Morrow, James D., Randolph M. Siverson and Tressa E. Tabares. "Correction to "the political determinants of international trade"." *American Political Science Review* 93.4 (1999): 931-933.

Noland, Marcus. Affinity and International Trade. Working Paper Series WP05-3, Peterson Institute for International Economics (2005).

Oneal, John R., Frances H. Oneal, Zeev Maoz, and Bruce Russett. The Liberal Peace: Interdependence, Democracy, and International Conflict, 1950 – 85. *Journal of Peace Research* 33.1 (1996):11–28.

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

Oneal, John R. and Bruce Russett. Triangulating Peace: Democracy, Interdependence, and International Organizations. New York, NY (2001): W. V. Norton.

Polachek, Solomon W.. Why democracies cooperate more and fight less: The relationship between international trade and cooperation. *Review of International Economics* 5. 3 (1997), 295–309.

Polachek, Solomon W. and Carlos Seiglie Trade, Peace and Democracy: An analysis of dyadic dispute. In Sandler Todd and Hartley Keith *Handbook of Defense Economics*, Vol. 2, Chapter 31 (2007): 1017-1073.

Powell, Emilia Justyna, and Stephanie J. Rickard. "International trade and domestic legal systems: examining the impact of Islamic Law." *International Interactions* 36.4 (2010): 335-362.

Przeworski, Adam, Michael E. Alvarez, Jose Antonio Cheibub, and Fernando Limongi. *Democracy and Development*. New York, NY (2000): Cambridge University Press.

Rose, Andrew K. "Do We Really Know That the WTO Increases Trade?" *American Economic Review* 94.1 (2004): 98-114.

Saikawa, Eri. "Policy diffusion of emission standards: Is there a race to the top?" *World Politics* 54.1 (2013): 1-33.

Simmons, Beth A. "Rules over real estate: Trade, territorial conflict, and international borders as institution." *Journal of Conflict Resolution* 49.6 (2005): 823-848.

Schumpeter, Joseph. Capitalism, Socialism, and Democracy. New York, NY (1950): Harper Torchbooks.

Souva, Mark, Dale L. Smith, and Shawn Rowan. "Promoting trade: The importance of market protecting institutions." *The Journal of Politics* 70.2 (2008): 383-392.

Tobin, Jennifer L. and Marc L. Busch, "A BIT is better than a lot: Bilateral Investment Treaties and Preferential Trade Agreements." *World Politics* 62.1 (2010): 1-42.

Appendix III. Additional Model Specifications and Clustered Standard Errors

In addition to posing challenges for hypothesis testing, dyadic data presents questions about model selection as well. Model selection involves the typical questions that arise with panel data. Should one incorporate fixed effects, and if so, which ones? Should one include a lagged dependent variable? The complicated structure of dyadic data exacerbates this debate. Indeed, model selection is a contentious issue in the democratic trade literature. (See Green, Kim, and Yoon 2001, Beck and Katz 2001, Oneal and Russett 2001, and King 2001).

Our main analysis in print includes fixed effects for dyad and year, which we see as the preferred specification. Others might disagree. In this appendix, we show that conventional significance tests are even more overconfident when the democratic trade hypothesis is modeled without fixed effects for dyad or year. This appendix also shows that clustering the standard errors by dyad is not a useful solution since regardless of the model, it too leads to overconfidence regardless of one's preferred model specification.

Table A.1 shows the regression results from six different models of democratic trade. In the first four columns are the results from a completely pooled model, a model with dyad fixed effects, a model with dynamics (a lagged dependent variable), and a model with dyad fixed effects and dynamics. In the last two columns are the results from the two models in our main text, one with dyad and year fixed effects and one with dyad and year fixed effects plus dynamics. In addition to the control variables in our main analysis, we include a control for the natural log of distance between the two countries in the dyad in the models with no fixed effects.

The conventional *p*-values and the randomization test *p*-values on the minimum democracy score within the dyad are displayed below the estimated democracy coefficients. Across all seven specifications, the conventional *p*-values are much smaller than the randomization test *p*-values, indicating that assumptions required for the parametric *t*-test are not met regardless of model choice.

Because the assumptions required for the conventional *t*-test do not hold, the test is quite overconfident and likely to commit a Type I error. Figure A.III.1 shows the distribution of the 1000 *p*-values associated with the conventional test calculated during each of the randomization test runs. The shaded areas cover *p*-values less than or equal to 0.1. If the test were appropriate for the data, we would expect to see false positives in the randomly shuffled dataset 10 percent of the time at the 90 percent confidence level. Instead, across each of the six different specifications, we find Type I error rates that are much too high, between 60 and 94 percent.

Still, our randomization test rejects the null hypothesis of no effect of democracy on trade in each of the specifications, albeit at more reasonable levels of confidence. Figure A.III.2 shows the randomization test results graphically. In each randomization test distribution, the observed test statistic is at least in the 5 percent most extreme statistics calculated on the randomly shuffled democracy scores.

These graphs also illustrate the compelling case for using fixed effects when employing this data. In theory, the distribution of test statistics obtained during the randomization should be centered at zero if there is no association between the variables. In other words, if the model is correct and if the time series of democracy scores are exchangeable, 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 each of the shuffled democracy score time series are random noise, one finds a negative relationship on average between democracy and trade in data with this structure. When dyad fixed effects are included without year fixed effects, the distributions are centered on positive values. Because democracy and trade both increase on average over time, differencing by dyad without accounting for this secular trend means that almost any random arrangement of democracy score will produce a positive coefficient. Again, this illustrates the importance of including both dyad and year fixed effects. Only when dyad and year fixed effects are included are the distributions centered at zero as they should be.

In conclusion, the extreme over-confidence we observed using the conventional *t*-test with our preferred model specification exists across all of the specifications, with or without dyad and year fixed effects. Thus it is clear that, regardless of one's preferred model, the often-employed parametric *t*-test should not be used with dyadic data like these.

The Limitations of Clustered Standard Errors

One common correction for the overconfidence of standard errors and hypothesis tests on clustered data is to estimate so-called cluster-robust (or clustered) standard errors. Cluster-robust standard errors allow for dependence between observations within the same cluster. Clustering the standard errors at the dyad level would account for any

unmeasured characteristics shared by every country A-country B dyad in the data over the years in which that dyad appears.

However, clustered standard errors do not allow for dependence between observations across clusters. Thus, one would have to assume, unrealistically, that causal forces (such as Nation A's degree of democracy) within the country A-country B dyad must be independent from those within all country A-country C dyads, even though country A is part of them all. Unfortunately, the cross-classified and multi-level structure of dyad-year data makes this particular fix unworkable, since one expects errors to be correlated within dyads, within nations, and over time.[4] While clustered standard errors in dyadic data would be an improvement over typical OLS standard errors, since clustered standard errors correct one type of dependence in dyadic data, they would still suffer from the problem of overconfidence.

For each of the models in Table A.1, we estimated cluster-robust standard errors clustered at the dyad level. The *p*-values from *t*-tests on minimum democracy score within the dyad using these adjusted standard errors are displayed in Table A.1, under the randomization test *p*-values. For each equation, the cluster-robust adjustment renders *p*-values that are still extremely small but larger than the typical *p*-values, confirming our intuition that there is clustering within dyads.

Across all of the models, the democracy coefficient remains highly statistically significant according to the cluster-robust *p*-values. Yet we should not trust these *p*-values because the disturbances are likely to be correlated both within dyads and also within each country in the dyad, violating a key assumption of the cluster-robust

[4] For two recent discussions of clustered standard errors in political science, see Beck (2012) and King and Roberts (2012).

estimators: that the errors are not correlated across clusters. Even the cluster-robust *p*-values are at least a million times smaller than the *p*-values obtained from our randomization tests, indicating that the robust standard errors are still overconfident. This confirms our intuition that the error structure is too complicated to be handled by this clustering algorithm.

Indeed, we can show using our simulations that cluster-robust standard errors do not fix the problem of over-confidence. In addition to recording the *p*-values from the *t*-tests on the 1000 randomly shuffled datasets, we also recorded the *p*-values from *t*-tests using clustered standard errors, clustered at the dyad level, for four of our six models. Figure A.III.3 shows that, while the type I error rates were lower for these tests, they were still much higher than the nominal confidence level would imply. Instead of a 10 percent false positive rate, the *t*-test with clustered standard errors falsely rejects the null hypothesis between 52 and 93 percent of the time. In short, clustered standard errors are not a viable solution to the complicated structure of dyadic data.

Works Cited

Beck, Nathaniel (2012). "Sweeping Fewer Things Under the Rug. Tis Often (Usually?) Better to Model than be Robust." Paper delivered at the 2012 Meetings of the Political Methodology Society, Chapel Hill, NC.

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.

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

King, Gary (2001). Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations. *International Organization* 55 (2): 497-507.

King, Gary, and Molly Rogers (2012). "How Robust Standard Errors Expose Methodological Problems They Do Not Fix." Paper delivered at the 2012 Meetings of the Political Methodology Society, Chapel Hill, NC.

Oneal, John R. and Bruce Russett (2001a). Clear and Clean: The Fixed Effects of the Liberal Peace. *International Organization* 55 (2): 469 – 485.

Figure A.III.1 Rate of False Positives in Conventional *T*-Test

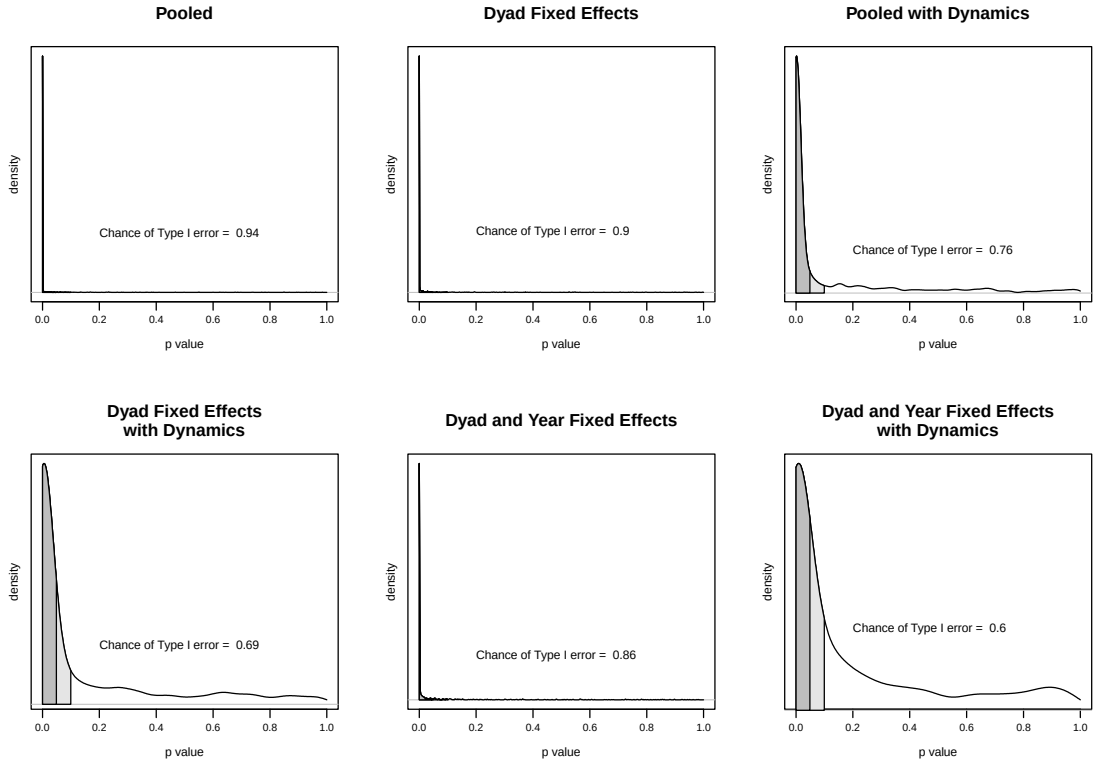

Figure A.III.1 shows the densities of the 1000 conventional *p*-values calculated during the randomization tests for the models in Table A.1. The shaded areas cover *p*-values that are 0.1 or smaller. The dark gray shaded areas cover *p*-values of .05 or smaller.

Figure A.III.2: Randomization Test Results

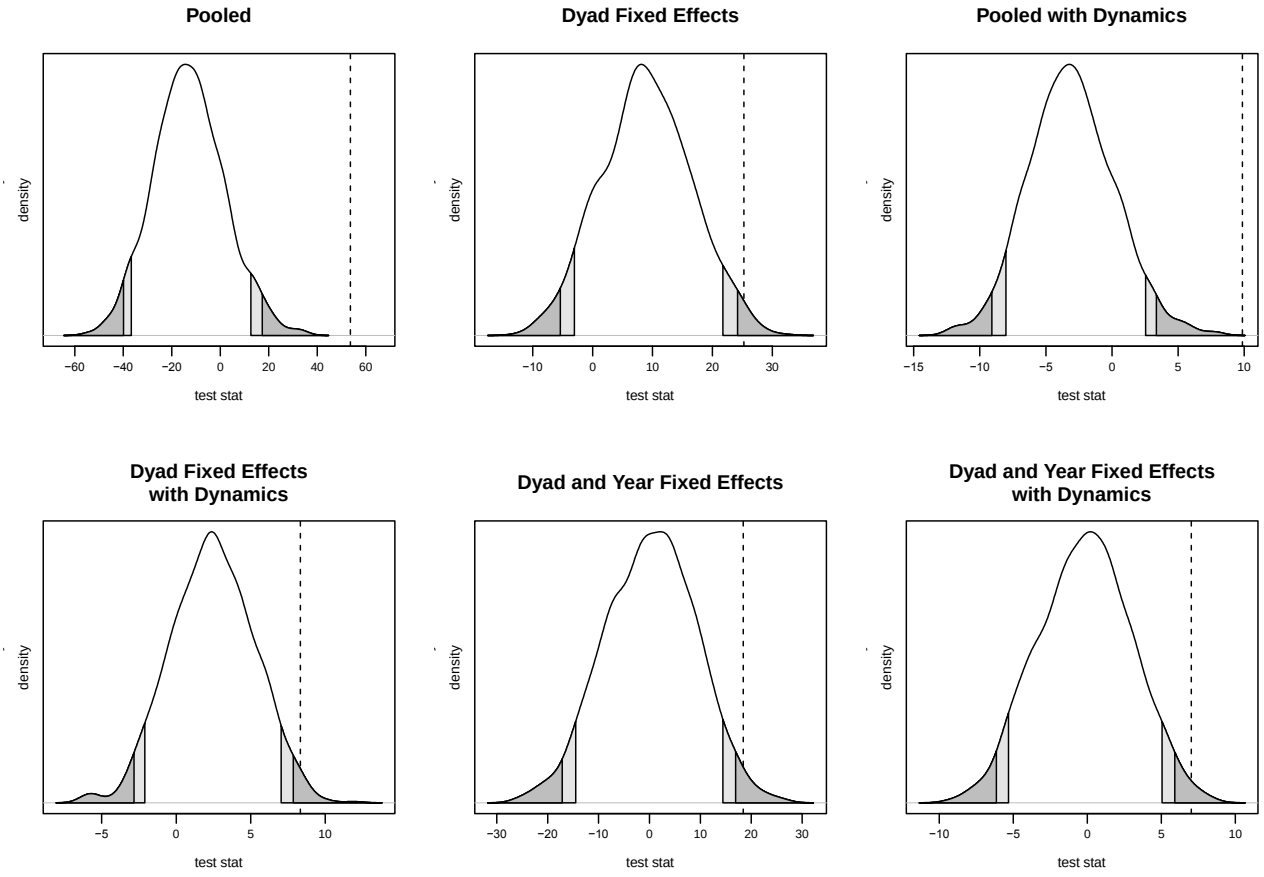

Figure A.III.2 shows the densities of the 1000 test statistics on the democracy variables estimated using the randomization test. The dark gray shaded areas represent the 5 percent most extreme test statistics, and the entire shaded areas cover the 10 percent most extreme test statistics. The dotted lines indicate the magnitudes of the test statistics estimated using the observed data, the ratio of the coefficients and standard errors in Table A.1.

Figure A.III.3: Rate of False Positives in *T*-Tests with Cluster-Robust Standard Errors

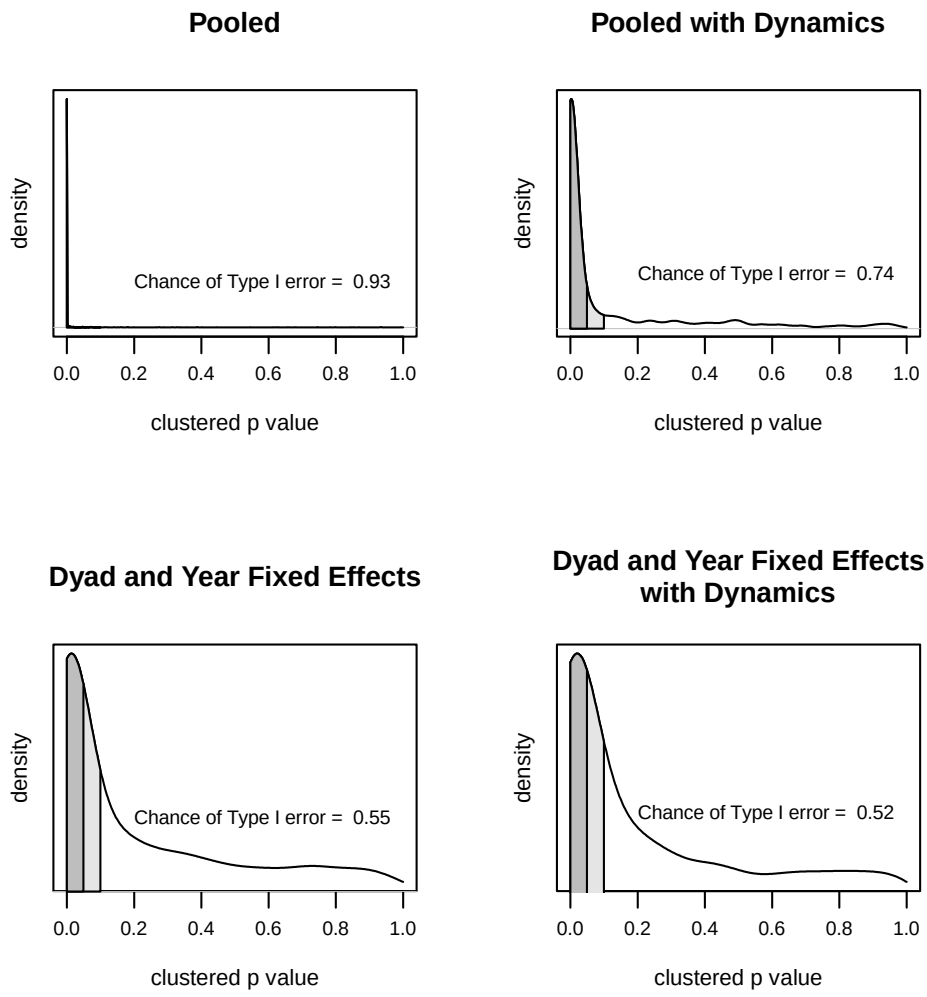

Figure A.III.3 shows the densities of the 1000 clustered *p*-values calculated during the randomization tests for the models in Table A.1. The shaded areas cover *p*-values that are 0.1 or smaller. The dark gray shaded areas cover *p*-values of .05 or smaller.

Table A1: Regression Analysis of Bilateral Trade, 1950-2000

	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	0.911 (0.003) *p=.000*	0.896 (0.005) *p=.000*	0.077 (0.002) *p=.000*	0.199 (0.004) *p=.000*	0.918 (0.007) *p=.000*	0.224 (0.005) *p=.000*
Population	0.592 (0.002) *p=.000*	0.190 (0.007) *p=.000*	0.052 (0.001) *p=.000*	0.040 (0.004) *p=.000*	0.252 (0.012) *p=.000*	0.103 (0.007) *p=.006*
Distance	-0.751 (0.006) *p=.000*		-0.063 (0.002) *p=.000*			
Alliance	0.351 (0.014) *p=.000*	0.378 (0.024) *p=.000*	0.028 (0.005) *p=.000*	0.065 (0.015) *p=.000*	0.384 (0.024) *p=.000*	0.064 (0.015) *p=.000*
Democracy	**0.039** **(0.001)** ***p<1.0x10^{-281*}*** ***rand p=.002*** ***clust p=7.7x10^{-31}***	**0.016** **(0.001)** ***p=3.7x10^{-140}*** ***rand p=0.011*** ***clust p=9.6x10^{-19}***	**0.003** **(0.000)** ***p=6.1x10^{-23}*** ***rand p=.014*** ***clust p=1.1x10^{-19}***	**0.003** **(0.000)** ***p=8.5x10^{-17}*** ***rand p=.013*** ***clust p=5.7x10^{-12}***	**0.013** **(0.001)** ***p=1.0x10^{-25}*** ***rand p=.036*** ***clust p=1.8x10^{-10}***	**0.003** **(0.000)** ***p=2.2x10^{-12}*** ***rand p=0.022*** ***clust p=3.8x10^{-9}***
Lagged bilateral trade			0.923 (0.001) *p=.000*	0.791 (0.002) *p=.000*		0.791 (0.002) *p=.000*
Constant	-17.580 (0.090) *p=.000*	-15.953 (0.097) *p=.000*	-1.498 (0.039) *p=.000*	-3.509 (0.067) *p=.000*	-17.312 (0.259) *p=.000*	-5.161 (0.178) *p=.000*
	N=119,640	N=2,346 T=51	N=117,294	N=2,346 T=50	N=2,346 T=51	N=2,346 T=50
R^2	0.66	0.87	0.95	0.95	0.87	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.
Except for those on Democracy, all $p < .0009$ are rounded to zero.
*This p value was too small for Stata or R to display.