Dirty Pool

Donald P. Green, Soo Yeon Kim, and David H. Yoon

The quantitative study of international relations is dominated by analyses of pooled cross-sections. When analyzing dependent variables, such as the occurrence of a militarized dispute or the level of trade between two nations, researchers tend to work with panel data sets of *NT* observations, where *N* is the number of dyads (pairs of nations) and *T* is the number of time points (typically years). Thus, for example, when sixty nations are observed annually over the span of forty years, the pooled cross-sectional data set consists of 1,770 dyads $\times$ forty years = 70,800 observations. These data are said to be "pooled" in that no distinction is made between observations in time and space. A datum is a datum, and one can draw inferences with equal certitude across dyads or across years.

Concerned that the effective number of observations is less than the nominal *NT*, a great deal of methodological attention has recently focused on problems of interdependencies among the observations; unobserved factors that cause the United States to go to war with Japan in 1941 also cause it to go to war with Italy and Germany. Nathaniel Beck and Jonathan N. Katz point out that ignoring these interdependencies may lead to biased inference if no corrections are made to the estimated standard errors associated with ordinary least squares (OLS) or probit.[1] While we share this methodological concern—as well as other concerns associated with the analysis of count data, sequential decisions, simultaneous equations, and

We are grateful to Neal Beck, Harry Bliss, Michael Boozer, Claudio Cioffi-Revilla, Robert Franzese, Jr., Geoffrey Garrett, Erik Gartzke, Alan Gerber, Simon Jackman, Gary King, Douglas Lemke, Jamie Morin, John Oneal, Bradley Palmquist, Frances Rosenbluth, Bruce Russett, Ian Shapiro, Alastair Smith, Allan Stam, William Thompson, James Vreeland, Michael Wallerstein, Stephen Walt, Erich Weede, and Nikos Zahariadis for comments and suggestions. We would like to thank the Institution for Social and Policy Studies at Yale, which supported this research. Data and computer programs used in this essay are available at the Web site http://web.polmeth.ufl.edu/research.html.

1. Beck and Katz 1995.

International Organization 55, 2, Spring 2001, pp. 441–468

rare events—we believe that the problems associated with standard pooled cross-sectional estimation run much deeper.[2]

We contend that analyses of pooled cross-section data that make no allowance for fixed unobserved differences between dyads often produce biased results. By "fixed unobserved differences" (or fixed effects, for short) we mean unmeasured predictors of the dependent variable that would cause each dyad to have its own base rate, or intercept. For example, year after year, trade levels between India and China fall below what one would expect based on a regression model that takes into account population size, gross domestic product (GDP), and shared borders. Because such a model fails to take note of the Himalayas, economic endowments, linguistic dissimilarity, and diplomatic relations, this model repeatedly overestimates bilateral trade between India and China, just as it consistently underestimates trade between Belgium and Switzerland. Pooling data implicitly assumes that the independent variables eliminate these persistent cross-sectional differences or render them uncorrelated with the predictors in the model. In this example, the fact that India-China differ in unmeasured ways from Belgium-Switzerland makes this assumption implausible. Given the vagaries of measurement and model specification in statistical studies of international relations, the statistical assumptions that underlie pooling are generally suspect.

In the next section, we describe in detail the strong econometric assumptions that are typically imposed when analysts perform pooled cross-sectional regressions in international relations. We make no attempt in this section to break new statistical ground; we merely summarize certain key issues that arise in the analysis of panel data and relegate some of the more technical discussion to the appendix. Turning next to the international relations literature, we survey dozens of recently published works in this area and find almost no attention to the problem of unmodeled fixed effects. To demonstrate the importance of this issue to students of international affairs, we present two empirical examples of how statistical results change when fixed effects are taken into account. The first example concerns bilateral trade; the second, militarized interstate disputes. In both cases, we find dramatic changes in the size and statistical significance of the parameter estimates. For example, democracy, which seems to be a leading predictor of peace in a pooled cross-sectional analysis, has no effect on militarized disputes when the data are examined longitudinally. We conclude by discussing the implications of our results for methodological practices in the field.

Pooled Cross-Sectional Models Versus Fixed-Effects Models

What is a pooled cross-sectional model, and how does it differ from a fixed-effects model? For simplicity of exposition, let us consider a linear regression analysis,

2. On count data, see Beck, Katz, and Tucker 1998; on sequential decisions, Signorino 1999; on simultaneous equations, Kim 1998; and on rare events, King and Zeng forthcoming.

deferring discussion of the complications introduced by binary dependent variables. The pooled cross-sectional model takes the form

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_K X_{Kit} + u_{it} \tag{1}$$

In this expression, the outcome Y_{it} is a function of K right-hand-side variables that vary across both time and space. The subscript i refers to one of the N cross-sectional units, and the subscript t refers to one of the T time points. The hallmark of this model is the inclusion of a *single intercept* (α) that reflects the expected value of the dependent variable when all of the independent variables are zero. To appreciate the practical implications of this assumption, imagine running this regression as a time-series analysis of a single dyad. In effect, this model makes the remarkable claim: "It doesn't matter which dyad one picks; the intercepts are all the same."

The pooled cross-sectional model in Equation (1) differs from a fixed-effects panel model in which each cross-sectional unit is assigned its own intercept. The regression model now includes $N - 1$ dummy variables, for each dyad (less one) in the data set:

$$Y_{it} = \alpha + \delta_1 Z_{1it} + \delta_2 Z_{2it} + \cdots + \delta_{N-1} Z_{N-1,it} + \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_K X_{Kit} + u_{it} \tag{2}$$

Here, the Z_{git} represent dummy variables marking each dyad, and the coefficients associated with each dyad are denoted δ_g. Thus the intercept for the first dyad is simply $\alpha + \delta_1$. Equation (1) is a subset of Equation (2), where all of the δ_g are constrained to be zero. Pooled cross-sectional regression, in other words, is a special case of a more general regression model.

Since pooled cross-sectional models omit variables that are included in the fixed-effects panel model, it should come as no surprise that pooled cross-sectional models may generate biased estimates of the β_k. When the δ_g are not zero (that is, when the dyads really do have different equilibrium levels) and when the Z_{git} are correlated with the X_{kit} (when the dyad-specific intercepts covary with the other independent variables), regression estimates will be biased. Note that these biases may be positive or negative, depending on how the intercepts covary with the regressors. In that sense, the problem of ignoring fixed effects is a special case of a more general problem, that of omitting variables in multivariate regression.

Scatter plots of hypothetical data illustrate how pooling may introduce bias. The plot in Figure 1 depicts a positive relationship between X and Y where $N = 2$ and $T = 50$. Because both dyads share a common intercept, pooling creates no estimation problems. One obtains similar regression estimates regardless of whether one controls for fixed effects by introducing a dummy variable for each dyad. A pooled regression is preferable in this instance because it saves a degree of freedom. In Figure 2 we encounter another instance in which pooling is benign. The two

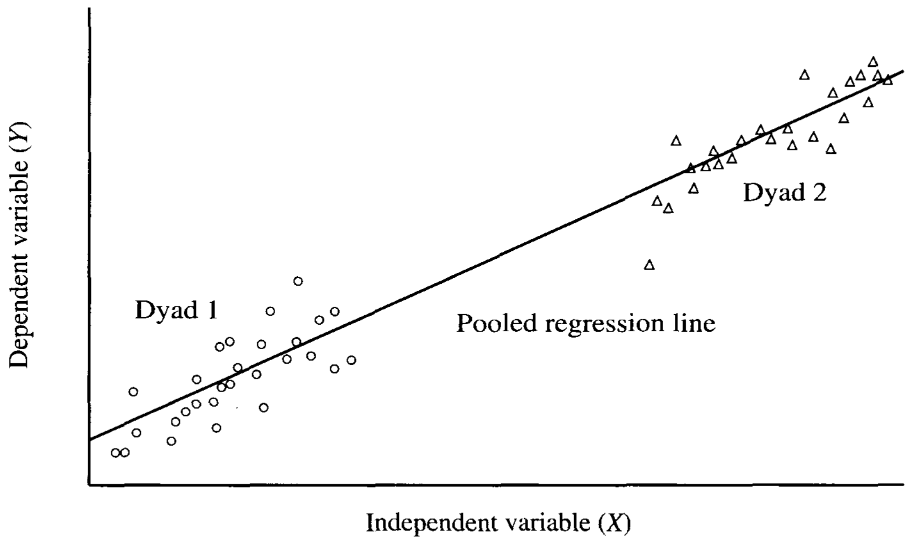

FIGURE 1. *Pooling homogenous observations*

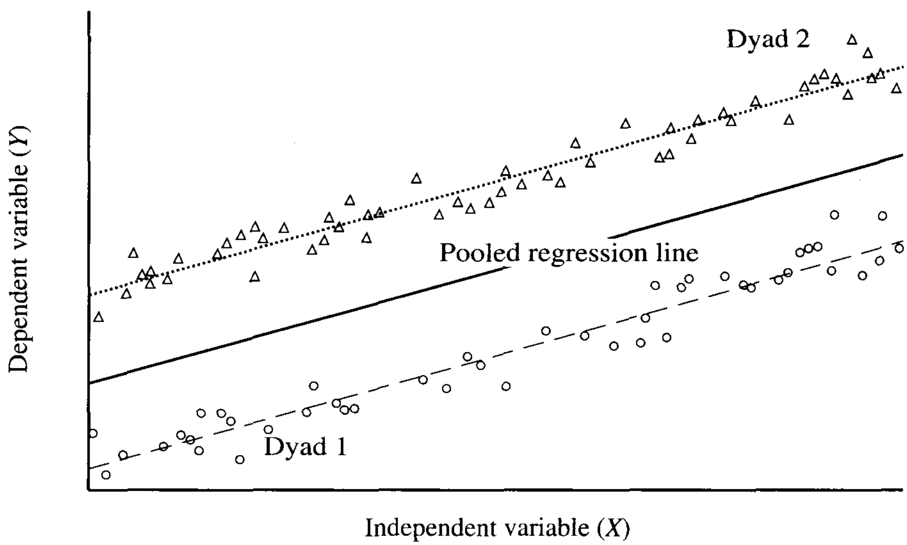

FIGURE 2. *Pooling dyads with randomly varying intercepts*

dyads have different intercepts, but there is no correlation between the intercepts and X. The average value of the independent variable is the same in each dyad. Again, pooled regression and regression with fixed effects give estimates with the same expected value. Figure 3 illustrates a situation in which pooled regression goes

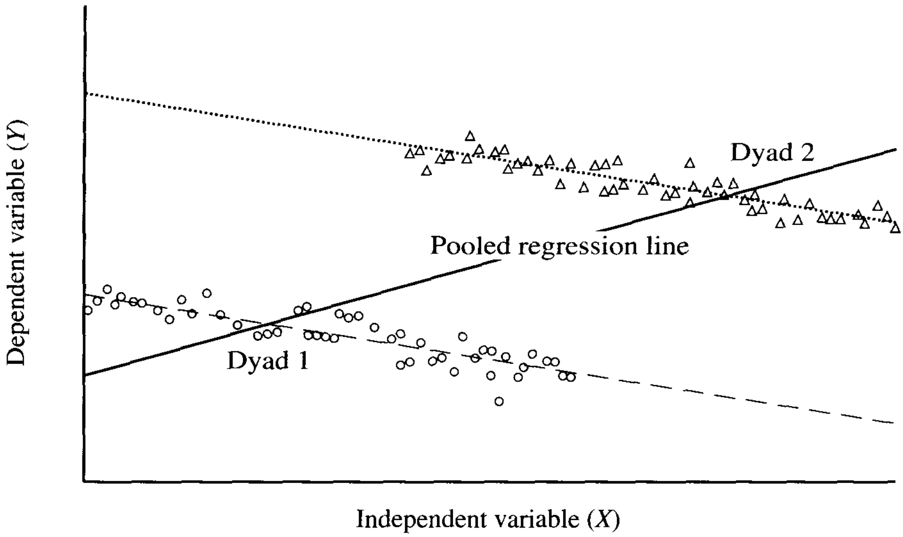

FIGURE 3. *Pooling observations ignoring fixed effects*

awry. Here, the causal relationship between X and Y is negative; in each dyad higher values of X coincide with lower values of Y. Yet when the dyads are pooled together, we obtain a spurious positive relationship. Because the dyad with higher values of X also has a higher intercept, ignoring fixed effects biases the regression line in the positive direction. Controlling for fixed effects, we correctly ascertain the true negative relationship between X and Y.

The foregoing examples suggest how fixed effects may be detected. By comparing alternative regression models, we obtain a sense of whether pooling causes estimation problems. This logic underlies statistical tests designed to check whether the assumptions behind Equation (1) are warranted. The most general null hypothesis is that all of the dyads have the same intercept ($\delta_g = 0$). For a linear regression, this amounts to an F-test with ($N - 1$, NT) degrees of freedom. Rejecting the null hypothesis suggests that the dyads have different intercepts, a necessary but not sufficient condition for bias. To establish bias, one compares the estimates of the β_k generated by Equations (1) and (2) using a Hausman test (see the appendix), which gauges whether the dyad-specific intercepts are random or instead correlated with the model's regressors (X_{kit}). Under the null hypothesis, dyads have distinctive intercepts, but these intercepts are uncorrelated with the regressors and therefore produce no bias. In this case, controlling for fixed effects produces *unbiased* but *inefficient* slope estimates; the slew of dummy variables wastes degrees of freedom, making the estimates less precise than they could be. The alternative hypothesis is that the dyad-specific intercepts are correlated with the regressors, in which case pooled regression will be biased and fixed-effects regression remains unbiased. The

Hausman test gauges whether the two sets of regression estimates differ to a significant degree. If so, the null hypothesis is false, and pooled regression is untrustworthy.

An interesting variant of Equations (1, 2) arises when one of the regressors is a lagged dependent variable. The pooled cross-sectional model can be changed to take into account time-series dynamics:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_K X_{Kit} + \beta_{K+1} Y_{i,t-1} + u_{it} \tag{3}$$

In this model, when $|\beta_{K+1}| < 1$, each dyad returns to an equilibrium value. Thus if the independent variables associated with two dyads take on the same values, those dyads will have the same equilibrium.

By contrast, when fixed effects are introduced, so that the underlying model is

$$Y_{it} = \alpha + \delta_1 Z_{1it} + \delta_2 Z_{2it} + \cdots + \delta_{N-1} Z_{N-1,it} + \beta_1 X_{1it} + \cdots + \beta_K X_{Kit} + \beta_{K+1} Y_{i,t-1} + u_{it} \tag{4}$$

each dyad may return to a different equilibrium level, even if the values of the X_{kit} are the same. Again, since Equations (3) and (4) are nested, one can test empirically whether the restrictions of the pooled cross-sectional model are sustainable.

The same methodological concerns apply to pooled cross-section analyses involving binary dependent variables, although additional complications arise as well. Instead of modeling levels in y_{it} as in Equation (1), the analyst models the probability that $y_{it} = 1$ (for example, war breaks out during a given year). In the case of logistic regression, the pooled model would look like this:

$$\text{Prob}(y_{it} = 1) = \frac{e^{a + B_1 X_{1it} + B_2 X_{2it} + \cdots + B_k X_{Kit}}}{1 + e^{a + B_1 X_{1it} + B_2 X_{2it} + \cdots + B_K X_{kit}}}$$

The corresponding fixed-effects model implicitly introduces dummy variables for each dyad rather than estimating a common intercept. It should be noted that in the fixed-effects specification, dyads that experience no variability over time in y_t have no effect on the likelihood function; they are in effect ignored by maximum likelihood estimators. This calls attention to an interesting conceptual point: Unless the analyst imposes special assumptions (such as a common intercept for all dyads), only those dyads that experience both war and peace can speak to the question of what causes war.[3]

3. We focus here on the consequences of ignoring fixed effects, but note that binary dependent variables introduce other complexities into dynamic modeling as well. Equations (3) and (4) apply to situations in which the dependent variable is continuous. When the dependent variable is limited, past realizations of that variable may contain considerable amounts of error. If a dyad has a .45 probability

The Use of Fixed Effects in Models of International Relations

Concern about the potential biases attendant to pooled cross-sectional regression is more widespread in some substantive fields than others. At one extreme, scholarship in the field of public finance features a preponderance of fixed-effects panel analyses and very few pooled cross-section regressions. Quite often, economists present results from a range of different panel models, a practice recommended by James A. Stimson.[4] At the other end of the continuum, scholars in international relations rely heavily on pooled cross-sections. This reliance is somewhat surprising given the frequent use of dyads as the unit of analysis. When nation-state is the unit of analysis, ten countries observed over twenty years give rise to two hundred observations; when converted to dyads, the N grows to 900. Dyads amplify the cross-sectional component of panel data. If fixed effects are thought appropriate for models of nations, one would think them even more appropriate for models of dyads.

Table 1 presents a methodological overview of recent panel analyses of militarized disputes, dyadic trade flows, and other phenomena related to international affairs or political economy. These articles, which appeared from 1996 to spring 1999, were culled from ten prominent journals in the fields of international relations and political economy: *American Journal of Political Science, American Political Science Review, Conflict Management and Peace Science, International Interactions, International Organization, International Studies Quarterly, Journal of Conflict Resolution, Journal of Peace Research, Journal of Politics,* and *Political Research Quarterly.* Table 1 lists these studies in reverse chronological order, sorting them into analyses that do or do not treat dyads as the unit of analysis. Among these fifty-one articles, the use of pooled cross-sectional models is well-nigh universal. Only three introduce fixed effects into the model. One of these, by Kurt Dassel and Eric Reinhardt, introduces fixed effects into an analysis of state-initiated violence, and, interestingly, this study did not treat dyads as the unit of analysis.[5] A few other studies come close to employing fixed effects, introducing dummy variables for each nation-state.[6] This practice, however, fails to capture any nonadditive features of pairs of nations. Across all dyads, it may be that the United States has an intercept of A and the United Kingdom an intercept of B, but if the intercept for the United States–United Kingdom dyad were not $A + B$, the one-nation-at-a-time method of controlling for fixed effects might prove inadequate.

It is possible that the practices of ignoring fixed effects or treating them as direct extensions of nation-specific effects reflect the nature of the subject matter. Perhaps

for war at time t, and we observe no war, the lagged dependent variable used to predict the probability of war at time $t + 1$ is scored zero.

4. Stimson 1985.

5. Dassel and Reinhardt 1999.

6. Mansfield and Bronson 1997.

TABLE 1. *Overview of recent quantitative analyses of panel data in international relations appearing in ten leading journals*

a. Analyses using dyads

Author(s) and year	*Dependent variable (years covered)*	*N*	*Method of analysis*
Leeds and Davis 1999	Dyadic interaction (1953–78)	35,578	OLS, general estimating equation, Huber-White standard errors
Wang 1999[a]	*UN voting coincidence rates (1984–93)*	*650*	*Least squares dummy variables (LSDV) regression*
Beck and Jackman 1998	Dyad-MID (1950–85)	20,990	Generalized additive model (GAM)
Beck, Katz, and Tucker 1998	Dyad-MID (1951–85)	20,990	Logit and grouped duration analysis
Bliss and Russett 1998	Bilateral trade (1962–89)	22,176	Generalized linear model with within-group correlation
Gartzke 1998	Dyad-MID (1950–85)	18,286–22,575	Logistic regression
Henderson 1998	Dyad-onset of war (1820–1989)	6,862	Logistic regression
Morrow, Siverson, and Tabares 1998	Bilateral trade (1907–90)	2,631	OLS, panel-corrected standard errors, Prais-Winston correction
Remmer 1998	Bilateral treaties (1974–85)	210	Logistic regression, negative binomial regression
Russett, Oneal, and Davis 1998	Dyad-IGO membership MID (1950–1985)	18,657–19,752	OLS, probit, robust standard errors
Bennett 1997	Length of time of rivalry (1816–1988)	423	Hazard model
Farber and Gowa 1997	Dyad-MID/war/alliance (1816–1980)	51,237–186,841	Probit
Mansfield and Bronson 1997	Bilateral trade (1960–90)	20,892–32,156	OLS, dummy variables for country and year, White standard errors
Mansfield and Snyder 1997	Dyad-war (1816–1986)	Not reported	Logit with(out) dummy variables for dyads with war
Mousseau 1997	Dyad-MID collaboration (1816–1992)	399,250	Logistic regression
Noland 1997[b]	U.S. trade policy behavior (1984–93)	370	Instrumental variables, probit, OLS
Oneal and Ray 1997	Dyad-MID involvement (1950–85)	6,355–20,990	Logistic regression, Huber-White standard errors
Oneal and Russett 1997	Dyad-MID (1950–85)	17,709–20,990	Logistic regression
Raknerud and Hegre 1997	Dyad-war (1840–1992)	965,166 (approx.)	Hazard model (Cox regression)
Thompson and Tucker 1997a	Dyad-alliance/war (1816–1913, 1946–76)	51,972; 163,578	Logit
Barbieri 1996	Dyad-MID; Dyad-war (1870–1938)	14,341	Logit
Bennett 1996	Rivalry termination/duration (1816–1998)	414	Logit, hazard model
Diehl, Reifschneider, and Hensel 1996	Dyad-recurring conflict (1946–88)	262 dyads–10 years	Event history analysis
Huth 1996	Enduring rivalry (1950–90)	3,039	Probit
Lemke and Werner 1996	Dyad-war (1820/1860–1980)	149–268	Logit
Oneal et al. 1996	Dyad-MID (1950–85)	6,641–22,575	Logistic regression
b. Analyses using states			
Bernhard and Leblang 1999	Exchange-rate arrangement (1974–95)	433	Constrained multinomial/binomial logit, period dummy variables
Blanton 1999	Human rights abuse (1982–92)	1,001	OLS, robust standard errors, lagged dependent variable
Dassel and Reinhardt 1999	*State-initiation of violence (1827–1982)*	*1,343–5,334*	*Fixed-effects duration dependent logit*
Keith 1999	Human rights behavior (1977–93)	2,149–2,478	OLS, panel-corrected standard errors, lagged dependent variable
Poe, Tate, and Camp 1999	Repression of human rights (1976–93)	2,144; 2,471	OLS, panel-corrected standard errors, lagged dependent variable

TABLE 1. *continued*

b. Analyses using states

Author(s) and year	*Dependent variable (years covered)*	*N*	*Method of analysis*
Apodaca and Stohl 1999	Foreign aid (1976–95)	252–1,424	Logit, OLS with presidential administration dummy variables
Benson and Kugler 1998	Severity of internal violence (1985–89)	130	Parks estimation (feasible GLS with autocorrelation)
Clark et al. 1998	Unemployment and output (1964–89)	1,682–1,765	OLS, lagged dependent and dummy variables, White standard errors
Enterline 1998	Dispute initiation (1816–1992, 1946–92)	10,579; 5,482	Logistic regression, natural cubic splines for duration dependence
Hall and Franzese 1998[c]	Inflation, unemployment (1955–90)	612	OLS, panel-corrected standard errors
Iversen 1998	Unemployment (1973–93)	75	OLS, lagged dependent variable, panel-corrected standard errors
Lemke and Reed 1998	Status quo evaluations	2,305–9,960	OLS, Granger-causality tests
Meernik, Krueger, and Poe 1998	U.S. foreign aid allocation (1977–1994)	1,475; 2,118	Probit, OLS, year dummy variables, robust covariance matrix
Perry and Robertson 1998	Political stochasticity (1955–1992)	57	OLS, period dummy variables, panel-corrected standard errors
Tir and Diehl 1998	State-MID (1930–1989)	4,801–5,734	Logit
Ward and Gleditsch 1998	State-war involvement (1815–1992)	10,681	Logit
Auvinen 1997	Domestic political conflict (1981–89)	56–630	OLS with(out) lagged dependent variable, one-way random effects, logit, tobit
Henderson 1997	Onset of interstate war (1820–1989)	2,535–4,817	Logistic regression
Leblang 1997	Restrictions on capital flows (1967–92)	538–2,157	Random effects probit
Leeds and Davis 1997	State's dispute behavior (1952–88)	666–2,664	Random effects probit, logistic regression with robust standard errors
Thompson and Tucker 1997b	State involvement in war (1816–1986)	7,571–8,929	Logistic regression, Huber-White standard errors
Zahariadis 1997	*State subsidies (1981–86)*	*48–54*	*Fixed effects model with(out) GLS for autocorrelation*
Hodgson 1996	Productivity growth (1870–1987)	144	OLS
Leblang 1996	Average per-capita growth rate (1960–90)	147	OLS, White standard errors, decade dummy variables
Simmons 1996	Money supply, bank rate adjustment (1925–38)	110–209	Two-way random effects (GLS)

Note: Studies using fixed effects are shown in italics.

Sources: The table summarizes studies appearing in the following journals since 1996: *American Political Science Review, American Journal of Political Science, Conflict Management and Peace Science, International Interactions, International Organization, International Studies Quarterly, Journal of Conflict Resolution, Journal of Peace Research, Journal of Politics,* and *Political Research Quarterly.*

[a]Wang analyzes the voting coincidence rate between the United States and sixty-five developing countries, as well as U.S. foreign aid to these countries. Thus although not explicitly stated in the tables, the dependent variable is essentially cast at the dyadic level.

[b]Noland's analysis of "the number of pages devoted to a country in the *National Trade Estimate Report on Foreign Trade Barriers* (ATTENTION)" (372) employs a fixed-effects estimator, though no details are given.

[c]Hall and Franzese also report estimation results using averages for the full 1955–90 period and also "decade" averages (1972–79, 1980–89) with weighted least squares and White standard errors.

different dyads are, other things being equal, similar in terms of their proclivities for trade or militarized disputes; perhaps the models and measures used in international relations adequately capture systematic variation across dyads. Thinking of international relations more generally, whether pooling "works" will depend on the nature of the dependent variable, how it is modeled, and the precision with which the predictors are measured. The only way to develop a sense of the robustness of pooled regression in international relations is to examine specific regression models and test the validity of pooling directly. To what extent do the coefficients of interest change when one controls for fixed effects? Granted, some coefficients will change simply because the fixed-effects estimator has more sampling variability than a pooled regression. The statistical question is whether the results change more profoundly than could be expected simply as a result of sampling variability.

Even if analysts hold strong priors in favor of the pooled model and the rigid assumptions it imposes concerning the unit-specific intercepts, they should at a minimum be concerned with the robustness of their findings across plausible alternative models.[7] As Stimson notes, pooled regression "may well be appropriate for a particular research question, but without entertaining other models there is no satisfactory way to know that it is appropriate."[8] We now examine the question of poolability empirically, showing how the substantive implications of large-*N* studies of trade and conflict change when one controls for fixed effects.

Data

Using a panel of dyads for the period 1951–92, we examine two dependent variables: bilateral trade volume and the presence or absence of a militarized interstate dispute.[9] For bilateral trade volume, the independent variables include the standard gravity model terms—log of *GDP*, *population*, and *distance* between capitals, and in addition, *alliance* and *democracy*. *Alliance* is operationalized as the absence (0) or presence (1) of a formal alliance; *democracy* as the lower of the net democracy scores within the dyad.[10] Trade data are from the *Direction of Trade Statistics* of the International Monetary Fund.[11] Data for GDP and population were obtained from the Penn World Tables, version 5.6.[12] The democracy variable was computed from the May 1996 version of the Polity III data set.[13] Data for contiguity,

7. Leamer and Leonard 1983.
8. Stimson 1985, 921.
9. See Bremer 1996; and Jones, Bremer, and Singer 1996.
10. For three large states (United States, USSR/Russia, and Canada), the shortest distance from their main ports/capitals is used. The ports include New Orleans and San Francisco for the United States, Vladivostok for USSR/Russia, and Vancouver for Canada. This measurement approach follows Bliss and Russett 1998; and Gowa and Mansfield 1993.
11. IMF 1997.
12. Heston and Summers 1991.
13. Jaggers and Gurr 1995.

capability ratio, and alliance were obtained from the Correlates of War Project (1995).[14]

The model of militarized interstate disputes features a set of commonly used regressors: *alliance, democracy, geographical contiguity*, the absence (0) or presence (1) of a shared land border; *capability ratio*, the ratio of the higher to lower capabilities indexes of the countries in the dyad, in logs; *growth*, the lower three-year average growth in per capita GDP within the dyad; and the lower *bilateral trade-to-GDP ratio* within the dyad.[15]

We have chosen to include in our analysis all dyads for which twenty or more observations were available, a criterion that admits over 93,000 of the approximately 117,000 cases in our data set. The reason for this restriction is that dynamic models are biased when estimated on short time-series. Note, however, that the coefficients we report are not changed appreciably when we admit all of the observations or, conversely, just those for which complete time-series data are available.[16]

In the interest of drawing an exact parallel between pooled-regression and fixed-effects regression, we include the same set of regressors in both models. Note that in the context of a fixed effects analysis, regressors such as contiguity and distance vary only insofar as countries divide or change their capitals over time. Just sixty dyads experience change in contiguity over time, but none experience change in distance. Distance is therefore a constant that is absorbed into the intercept associated with each dyad. Fixed-effects regression turns a blind eye to such time-invariant regressors; to learn about their effects, one must either study them in a cross-sectional context, braving the usual threats to causal inference, or investigate particular historical instances in which observations vary over time.

Results

We begin our panel analysis by modeling a continuous dependent variable, the total volume of trade between two states (in logs). Our specification includes the three components of the "gravity model"—log of the two states' total GDP, the log of the two states' total population, and the log of the distance between the two states.[17] As Jeffrey H. Bergstrand cautions, this model offers reasonably accurate predictions of

14. Singer and Small 1994.

15. For a summary of the capability index, see Singer 1990.

16. In an earlier draft of this article, we reported results from a "balanced panel," which is a panel restricted to just those dyads with complete data for the entire time span (1961–89). The coefficients were similar to those reported here, but the loss of observations made for larger standard errors. Despite a sample of more than 29,000 observations, no predictors of militarized disputes were significant at the 5 percent level in a regression that controlled for fixed effects.

17. See Tinbergen 1962; Linneman 1966; Leamer and Stern 1970, 145–70; Anderson 1979; and Deardorff 1984, 503–504.

TABLE 2. *Alternative regression analyses of bilateral trade (1951–92)*

Variable[a]	*Pooled*	*Fixed effects*	*Pooled with dynamics*	*Fixed effects with dynamics*
GDP	1.182**	0.810**	0.250**	0.342**
	(0.008)	(0.015)	(0.006)	(0.013)
Population	−0.386**	0.752**	−0.059**	0.143*
	(0.010)	(0.082)	(0.006)	(0.068)
Distance	−1.342**	Dropped: no within-group variation	−0.328**	Dropped: no within-group variation
	(0.018)		(0.012)	
Alliance	−0.745**	0.777**	−0.247**	0.419**
	(0.042)	(0.136)	(0.027)	(0.121)
Democracy[b]	0.075**	−0.039**	0.022**	−0.009**
	(0.002)	(0.003)	(0.001)	(0.002)
Lagged bilateral trade			0.736**	0.533**
			(0.002)	(0.003)
Constant	−17.331**	−47.994**	−3.046**	−13.745**
	(0.265)	(1.999)	(0.177)	(1.676)
	N = 93,924	NT = 93,924	N = 88,946	NT = 88,946
		N = 3,079		N = 3,079
		$T \geq 20$		$T \geq 20$
Adjusted R^2	0.36	0.63	0.73	0.76

Note: Estimates obtained using *areg* and *xtreg* procedures in STATA, version 6.0.
[a]GDP, population, distance, and bilateral trade are natural-log transformed. Method of analysis is OLS and fixed-effects regression.
[b]Lower value within the dyad.
**$p < .01$.
*$p < .05$, two-tailed test.

trade volume but lacks firm theoretical foundation.[18] Political scientists have treated the gravity model as something of a baseline, appending additional political variables. We follow current practice in the spirit of examining the consequences of different modeling assumptions. We include as regressors the democracy and alliance measures from the previous analysis. Table 2 presents both pooled and fixed-effects models, each with and without a lagged dependent variable as a regressor. We find no support whatsoever for the null hypothesis that all dyads share the same intercept. For the nondynamic case, $F(3078,90841) = 23.68$, $p < .0001$; when lagged trade is introduced as an independent variable, $F(3078, 85862) = 4.43$, $p < .0001$.

Clearly, the dyads have different intercepts, but are these omitted intercepts a source of bias for pooled regression? The correlation between the dyad-specific

18. Bergstrand 1985, 474.

intercepts and the predicted values derived from the fixed effects model is .42, suggesting that pooling is deeply problematic. A Hausman test establishes this point more rigorously by comparing the fixed-effects estimates with those derived from a model in which intercepts are presumed to be uncorrelated with the regressors. The two sets of regression estimates are found to be significantly different, and we decisively reject the so-called random-effects model (of which the pooled regression is a special case): $\chi^2(4) = 870.0$, $p < .0001$. Pooled regression is biased.

With large data sets it is sometimes possible to reject parsimonious regression models in favor of somewhat more complex models that produce substantively identical results. That is manifestly not the case here. The two regressions paint markedly different pictures of bilateral trade. In the pooled analysis, population has a strong negative effect on trade. A one-unit change in the log of population *reduces* the log of trade by .39 units. The tiny standard error associated with this estimate produces a T-ratio of epic proportions, –39.7. Not in a million years could these data have been generated by a true parameter of zero or more. Yet, look at the fixed-effects regression results: population has a positive coefficient (.75) and a T-ratio of 9.2. As two countries' populations grow over time, other things being equal, they trade more.[19] Alliance and democracy undergo similar turnabouts. In the pooled model, democracy encourages trade. In the fixed-effects model, dyads trade less as the less-democratic partner becomes more democratic. In the pooled model, alliance inhibits trade. In the fixed-effects model, the formation of alliances is associated with much higher levels of trade.

Similar turnabouts occur when we introduce a lagged dependent variable and focus on the short-term influences of the independent variables. Again the Hausman test indicates that the pooled cross-sectional regression is biased (a test against a null hypothesis of random effects produces $\chi^2(5) = 14{,}754.0$, $p < .0001$), and we see dramatic changes in the magnitude of the slope estimates associated with population, alliance, and democracy. As expected, the pooled model overestimates the effect of the lagged dependent variable. The coefficient that the pooled model assigns to the lagged dependent variable blends the true parameter with the parameter of unity that should be assigned to its (omitted) intercept. Because the effect of the lagged dependent variable is overestimated, it appears that perturbations to trade levels reequilibrate more slowly than they actually do.[20] In sum,

19. As noted earlier, the causal interpretation of coefficients growing out of the gravity model is problematic. Leamer and Stern (1970, 155) argue persuasively that population change may reflect a variety of unmeasured variables, such as technological change and changing health care. Note also that the gravity model makes no distinction between imports and exports, which might be differentially affected by trade volume. For these reasons, we are loath to say what constitutes the "right" sign for the population coefficient.

20. Results similar to the fixed-effects regression obtain when we use an alternative estimator that makes allowance for the fact that lagged trade is an endogenous regressor. This alternative estimator uses the Anderson-Hsiao methodology (instrumental variables) described in Hsiao 1986 and Greene 1997. These results are available from the authors on request.

TABLE 3. *Alternative logistic regression analyses of militarized interstate disputes (1951–92)*

Variable	*Pooled*	*Fixed effects*	*Pooled with dynamics*	*Fixed effects with dynamics*
Contiguity	3.042**	1.902**	1.992**	1.590**
	(0.092)	(0.336)	(0.120)	(0.375)
Capability ratio	0.102**	0.387**	0.125**	0.350*
(log)	(0.024)	(0.139)	(0.028)	(0.151)
Growth[a]	−0.017	−0.059**	−0.026*	−0.062**
	(0.011)	(0.012)	(0.013)	(0.013)
Alliance	−0.234*	−1.066*	−0.013	−1.090*
	(0.097)	(0.426)	(0.118)	(0.526)
Democracy[a]	−0.057**	−0.003	−0.053**	0.0004
	(0.007)	(0.015)	(0.008)	(0.016)
Bilateral	−0.194*	−0.072	0.028	0.084
trade/GDP[a]	(0.087)	(0.186)	(0.075)	(0.217)
Lagged dispute			4.940**	1.813**
			(0.102)	(0.103)
Constant	−5.809**		−6.274**	
	(0.090)		(0.108)	
N	93,755	93,755[b]	88,752	88,752[c]
Log likelihood	−3,688.06	−1,546.53	−2,530.31	−1,299.53
χ^2	1,186.43	75.75	3,074.67	380.40
Degrees of freedom	6	6	7	7
Prob > χ^2	<0.0001	<0.0001	<0.0001	<0.0001

Note: Estimates obtained using *logit* and *clogit* procedures in STATA, version 6.0.
[a]Lower value within the dyad. Method of analysis: Logistic and fixed-effects logistic regression.
[b]2,877 groups (87,402 observations) have no variation in outcomes.
[c]2,883 groups (82,932 observations) have no variation in outcomes.
**$p < .01$.
*$p < .05$, two-tailed test.

assumptions implicit in different regression models greatly shape how one thinks about the determinants of bilateral trade.

To illustrate further the importance of fixed effects, we turn our attention to a nonlinear estimation problem. Table 3 reports the results of alternative logistic-regression models of militarized disputes. The pooled analysis suggests that the likelihood of disputes increases when dyads are contiguous and decreases as the less-democratic member of the dyad becomes more democratic. Alliances decrease the risk of war, whereas differences in military capabilities increase it. These results are in line with published research.

These estimates change markedly when fixed effects are controlled. Democracy's effects become negligible and statistically insignificant, whereas military capability and alliance prove much more influential. Consider, for example, what the fixed-effects regression results tell us about a dyad with a 5 percent chance of war. If the less democratic of the two nations becomes fifteen units more democratic, the risk of war decreases to 4.8 percent. The pooled regression would lead us to expect this risk to drop from 5 percent to 2.2 percent. Conversely, the formation of an alliance decreases the risk of war from 5 percent to 1.8 percent, not 4.0 percent, as implied by the pooled regression. Controlling for fixed effects changes the way one views the relative importance of regime type, bilateral accords, and military capabilities.

Another important insight to emerge from this modeling exercise concerns the size of the standard errors. Because fixed-effects regression uses a great many degrees of freedom and eliminates cross-sectional variation in the independent variables, the standard errors are sometimes several times larger than those generated by pooled cross-sectional analysis. It turns out that only 198 of the dyads have any longitudinal variation in the dependent variable. The remaining 2,877 dyads are all zeros and have no effect on the regression results. Whether this fact constitutes an appalling waste of data or a necessary protection from bias depends on the tenability of the assumptions underlying pooling. Here, a Hausman specification test decisively rejects the null hypothesis that pooled regression is unbiased ($\chi^2(6) = 190.2$, $p < .0001$), and we reject as well the hypothesis of random effects ($\chi^2(6) = 73.2$, $p < .0001$).

Note that if fixed-effects regression merely left us awash in uncertainty due to lack of variation over time, the Hausman test would fail to reject the null. *Fixed-effects regression shoulders the burden of proof; a dearth of time-series information would lead us to accept the adequacy of pooling.* These lopsided Hausman tests show that the time-series information is indeed sufficient to expose the defects of pooling. Pooled cross-sectional analysis of militarized disputes produces misleading results.

At first blush, the fixed-effects specification seems open to the charge of selection bias. It would appear that cases are discarded according to the values of the dependent variable. It should be stressed that dyads without temporal variation are not "dropped" in the ordinary sense of the term. In regression analysis, dropping cases changes the way that the regression estimates are computed. In conditional logit, the time-invariant observations simply add zero to the likelihood function and are computationally irrelevant. This is more than a bit of statistical trivia. It means that variation in the dependent variable is a necessary condition for a dyad to be informative within the context of a fixed-effects model. Consider the hypothetical example of an as-yet undiscovered nation (perhaps on a deserted island or on another planet). One day, this nation is discovered, and international relations scholars document its record of unbroken peace with each of the known nations. Has anything been learned about the causes of militarized disputes from the inclusion of these new cases? No, because we do not know the base probability (the intercept) of war for each of these new dyads, parameters that cannot be identified without some variation in the binary dependent variable or some theoretically driven stipulations.

Need we model each dyad's intercept, or can we get by with a more parsimonious accounting of cross-dyad variation? One plausible approach is to introduce intercepts for each of the ninety-six countries rather than each of the 3,075 dyads analyzed in column 2 of Table 3, as in the study by Edward D. Mansfield and Rachel Bronson.[21] The null model maintains that, controlling for country-level effects, any dyad-specific variation in intercepts is uncorrelated with the regressors. A Hausman test soundly rejects this null hypothesis ($\chi^2(6) = 79.9, p < .0001$). Returning to the bilateral trade data, we find the same thing. Country-level effects in no way solve the problem of bias ($\chi^2(4) = 508.7, p < .0001$). Evidently, dyad-specific fixed effects are not simply reducible to country-level effects; including only the latter produces misleading results. And since country is a lower level of aggregation than region, the same conclusion holds for pooled cross-sectional analysis within regions. Parsimony has its allure, but if one is to impose a more parsimonious scheme of fixed effects on these data, it must be more nuanced than country or region. Just what such an arrangement of fixed effects might look like in practice is by no means obvious.

Indeed, it could be argued that we have not gone far enough in our search for sources of bias. Rather than simply account for fixed effects associated with each dyad, we could also have accounted for fixed effects associated with each year. In the interest of brevity, we do not present those results here. Suffice it to say that for both militarized disputes and bilateral trade, fixed effects associated with each year significantly improve each model's fit but do not greatly alter the substantive implications of the fixed-effects regressions presented in Tables 2 and 3. Exploring a range of potential fixed-effects specifications need not draw the analyst into a vortex of ever-changing results.

It should be stressed that the point of this exercise is to demonstrate that the pooled models are sensitive to subtle assumptions about the ways in which dyads are modeled over time and space.[22] Coefficients change, and standard errors sometimes increase. Although these changes are in some cases startling, we resist the temptation to draw any particular substantive conclusions about the sources of militarized disputes or international trade because the fixed-effects models that we present in no way resolve a range of nagging methodological problems arising from reciprocal causation, inadequate measurement, selection bias, and the like. Our thesis is not that fixed-effects regression solves the methodological problems that bedevil research on international trade and security. The investigation of fixed effects is part of a broader inquiry into the validity of parameter restrictions across time and space. We have examined whether dyads have different intercepts and alluded to the possibility of systemwide shocks occurring at particular points in time. One might also ask whether dyad-level intercepts remain constant over time, as factors outside the model cause a dyad's intercept to drift. More generally, one may ask whether slopes vary across dyads or over time, raising broader questions about the trans-

21. Mansfield and Bronson 1997.

22. See Stimson 1985. One such assumption is that the slope parameters are constant across dyads and over time. One could argue that unmodeled interactions exist either among the independent variables or between these variables and omitted regressors, as in Ragin 1987. The introduction of fixed effects should be viewed as a first step in the direction of testing and perhaps relaxing these constraints.

portability of models across historical periods. How far scholars must go in this direction depends both on the nature of the subject matter and the quality of the data used to study it. Some dependent variables may prove recalcitrant, showing signs of fixed effects in time or space or both. In other cases, pooled regression may be unproblematic.

Having cautioned the reader against placing undue faith in fixed-effects models, we nonetheless believe that testing for fixed effects will put the quantitative analysis of panel data in international affairs on a path toward more robust and informative models. Many years after Stimson's watershed essay on the analysis of panel data in political science, it seems clear that the assumptions underlying pooled cross-sectional analysis of trade and conflict are suspect.[23] Dyads differ systematically in ways that are not captured by the measures used to gauge constructs such as "capability," "democracy," and the like. Pooling data under these circumstances leads to biased estimates. Yet analysts of international relations seem unaware of this problem or unwilling to come to grips with it.

The persistence of fixed effects in the cases examined here should be seen as a challenge to future scholarship on trade or interstate disputes: find new regressors that capture these cross-sectional differences. As Stimson points out, fixed effects are merely placeholders awaiting substantive explanation.[24] Scholars rising to this challenge may then judge their handiwork according to whether their revised regression models succeed in transforming any remaining dyad-specific intercepts into random noise, as gauged by a Hausman test. This approach, if successful, could resuscitate the cross-sectional component of panel analysis and turn pooled regressions into the kinds of conditional random-effects models envisioned by Simon Jackman.[25] Until then, analysts of pooled cross-sectional data should proceed with caution, and consumers of this research should begin to demand that scholars consider potential problems arising from unmodeled fixed effects.

Discussion

In closing, we wish to address three likely rejoinders to our critique. The first is the notion that fixed-effects regression eliminates the most "interesting" part of the variance, namely, cross-sectional variance. The comparative method naturally impels scholars to compare different dyads, not just track them over time. Although this critique has a certain rhetorical appeal, it only makes sense from a mathematical standpoint if one accepts the strong homogeneity assumptions underlying the pooled cross-section models in Equations (1) or (3). Cross-sectional variation reflects both modeled and unmodeled differences among dyads. One evades the problem of bias

23. Stimson 1985. It is noteworthy that Stimson is cited by Wang 1999 and Zahariadis 1997, two of the three works listed in Table 1 that introduce fixed effects into their international relations models.
24. Stimson 1985.
25. Jackman 1999.

only by assuming that one's model accounts for all cross-sectional variability such that any remaining differences across dyads have nothing to do with the included regressors. Whether cross-sectional variance is interesting or misleading hinges on the validity of this assumption. Statistical tests have shown this key assumption to be empirically unsupported for bilateral trade and militarized disputes. In time, students of international relations will discover whether similar problems attend to dependent variables such as the formation of alliances, arms transfers, and the like.

It is both telling and troubling that those who pool over time and space frequently use the language of temporal dynamics when illustrating the causal implications of their statistical results. As a country becomes more democratic, the dyads of which it is a part are said to become more peaceful.[26] When two countries enter into an alliance, their level of trade is thought to decline.[27] As Stanley Lieberson points out, temporal connotations of "change" may not follow from cross-sectional comparisons, because the latter may give biased indications of cause and effect.[28] That point seems especially appropriate here, for we have seen that important results from pooled cross-sectional models may be driven entirely by cross-sectional variability. When dyads are traced over time, the substantive implications may look altogether different. Cross-sectional inference is not inherently invalid, but it cannot be considered reliable if contradicted by time-series analysis. At a minimum, scholars must investigate why these two analytic approaches give different answers. Our reading of recent studies that use pooled regression suggests that few researchers have grappled with the possible tension between the cross-sectional and cross-temporal components of their analyses.

A second argument concedes that the pooled regression results are driven largely by cross-sectional variation but contends that cross-sectional analysis provides a valuable means of studying which countries are likely to go to war. This defense of cross-sectional inference distinguishes between two kinds of research objectives: if we want to know when countries will fight, it makes sense to track dyads over time; but if we want to know who will fight, then pooling without fixed effects makes sense. This argument subtly confuses parameter estimation with prediction. If we are only interested in predicting (or describing) which countries go to war, then any model, including the use of pooled regression, is fair game. Democratic countries are less likely to fight each other; allied countries trade less. Such statements are agnostic about the causal questions of whether democratization reduces the risk of war or whether entering into alliances reduces trade.

If, however, we are interested in estimating the structural parameters that govern cause and effect, as in Equation (2), *time-series analysis and cross-sectional analysis should, in principle, give the same answers*. The attraction of cross-

26. See examples of "changing" levels of democracy in Oneal and Ray 1997, 767, tab. 4; and Mousseau 1997, 83, fig. 2.

27. See Morrow, Siverson, and Taberes 1998, 659, tab. 3; and Mansfield and Bronson 1997, 100, tab. 2.

28. Lieberson 1985, chap. 9.

sectional analysis is that it affords us the opportunity to study large numbers of cases; the drawback is that estimates will be biased if the model omits variables that are correlated with the independent variables. The attraction of time-series analysis is that it may reduce the risk of spurious comparisons; the drawback is that one typically has fewer discrete observations of time-units for which reliable time-varying measurements are available.[29] Panel analysis with fixed effects blends the two approaches, garnering large numbers of observations while estimating parameters within units of observation. Panel analysis is no panacea. The independent variables in the model may be mismeasured or correlated with omitted variables that change over time, in which case the results will be biased. The hope is that by examining an assortment of dyads that change in different ways over time, the risk of bias is reduced.

Again, we would stress that cross-national comparison is not inherently invalid; it is simply risky, insofar as it involves strong assumptions about the nature of cross-sectional variability. Granted, if these strong assumptions are satisfied, pooled regression is more efficient than fixed-effects regression, but the question is whether these assumptions are satisfied. By offering a means by which to corroborate or contradict cross-sectional results, panel data help to illuminate the validity of these assumptions. Using the Hausman tests described earlier, it is a relatively simple matter to see whether the allure of efficiency outweighs the threat of bias.

It might seem that we are holding analysts of pooled cross-sectional data to a higher standard than analysts of cross-sectional data, but the same logic applies to both even if the opportunities for corroboration are more readily available to those using panel data. Cross-sectional findings direct our attention to out-of-sample predictions. In our data set, for example, the variable "distance between countries" is constant over time, so our knowledge of its effects is based entirely on cross-sectional inference, which would be misleading if distance were correlated with omitted factors such as ethnic antagonism. Looking beyond the bounds of our data set, we might consider instances where buffer states emerge between formerly contiguous countries, asking whether their hostility and trade subsequently diminishes, as cross-sectional analyses suggest.[30]

The example of distance brings us to the final rejoinder, which is an appeal to necessity. The argument goes like this: If one eliminates all time-invariant regressors and cross-sectional variation in intercepts, variables such as regime type lose almost all of their variance, especially if the analysis is restricted to a particular

29. Obviously, one could have arbitrarily larger T if the unit of observation were, say, "dyad hours" instead of dyad years, but one does not possess measures of the independent and dependent variables at such fine time gradations; Beck and Tucker 1997. Problems of autocorrelation among the disturbances (causative factors omitted at one point in time are omitted in the next) become more pressing as the time unit becomes shorter. Autocorrelation biases the standard errors generated by pooled regression.

30. An example of a cross-sectional analysis bolstered by comparisons over time may be found in Russett 1978. Russett finds a strong negative cross-national relationship between per capita GNP and infant mortality and corroborates this result by tracing select countries over time.

region or time period.[31] If we cannot infer the effects of regime type from comparisons of different dyads (United States–United Kingdom versus United States–China), it will be impossible to draw any reliable inferences at all. Pooled cross-sectional analysis may be biased and may unjustifiably duplicate N observations T times, but it is all we can do given the dearth of variability over time.

It is entirely possible to raise questions that cannot be answered reliably with available data. One approach, as we noted earlier, is to develop new independent variables that render the pooled regression model of Equation (1) more palatable. Alternatively, one could bring to bear prior knowledge that a select group of dyads has distinctive intercepts, excluding certain fixed effects and allowing some cross-sectional variation to influence the estimates (see the appendix). Typically, however, the data analyst lacks a ready reserve of compelling independent variables or strong prior convictions about which fixed effects can be ignored. Under these circumstances, it is sensible to defer to statistical tests that gauge the adequacy of the pooled-regression model (and related models that assume the dyad-specific intercepts to be unrelated to the independent variables). When such specification tests show these models to be biased, fixed-effects regression is an appropriate corrective. The fixed-effects regression estimates may be accompanied by large standard errors, but that is an indication of the uncertainty associated with cross-sectional comparisons. Since the alternative in this case is a misleading pooled analysis, fixed-effects regression can scarcely be faulted for being the bearer of bad tidings. Drawing secure inferences about causality in international affairs is an extraordinarily difficult undertaking. Pooled regression merely pretends that this is not so.

Appendix

Hausman's Specification Test

If θ^{eff} denotes an estimator that is asymptotically efficient under a null hypothesis but inconsistent when the null is false and θ^{cons} denotes an estimator that is inefficient under the null but remains consistent under the alternate, Jerry Hausman[32] shows that a chi-square test statistic can be constructed such that

$$\chi^2[k] = (\theta^{cons} - \theta^{eff})'[Var(\theta^{cons}) - Var(\theta^{eff})]^{-1}(\theta^{cons} - \theta^{eff}), \tag{5}$$

where k, the degrees of freedom, is equal to the dimension of the vector θ. The intuition for the test is that if the null is true, the difference between the two estimators, $\theta^{cons} - \theta^{eff}$, should be asymptotically small since both are consistent

31. Huber, Ragin, and Stephens 1993, 733, n.19
32. Hausman 1978.

under the null. With the appropriate weight matrix, the quadratic form in the difference has an asymptotic chi-square distribution.[33]

Heterogeneity and Fixed Versus Random Effects

Given a panel data set of N individuals (in our case, dyads) and T yearly observations for each individual, let y_{it} be the dependent variable (magnitude of militarized disputes or trade between states in dyad i) in year t. And let X_{it} be the vector of explanatory variables for dyad i in year t, with α_i representing the unobservable effect for individual i. For a linear specification, we can write this general model as

$$y_{it} = X_{it}\beta + (\alpha_i + u_{it}), \quad (i = 1,\ldots,N;\ t = 1,\ldots,T). \tag{6}$$

Note that this is an equivalent reformulation of the linear model presented in the body of the article.

A commonly encountered presumption in regression analyses of panel data in international relations is that $\alpha_i = \alpha$ for all i, or, more specifically, that unobserved dyad-specific effects are nonexistent or unimportant. But a priori pooling based on an assumption of homogeneity is tantamount to ignoring the principal structure of panel data—repeated observations of the same set of individuals or dyads. Presumably, observations from the same dyad will be more "alike" than observations from different dyads.[34] Whether this is indeed the case must be tested.[35] In the bilateral trade model, we use an F-test to reject the hypothesis of homogeneity.[36]

Once heterogeneity is suspected, the individual effects can be modeled as either *fixed* or *random*. The fixed-effects approach, on the one hand, assumes that the heterogeneity is equivalent to parametric shifts in the model specification: α_i in Equation (6) would be the individual-specific constant term representing such shifts. The random-effects approach, on the other hand, assumes that the heterogeneity is due to individual-specific components of the disturbance. An equivalent statement of the distinction is that the fixed-effects model allows for correlation between the α_i's and the columns of X_{it}, whereas the random-effects model assumes that the α_i's are uncorrelated with the columns of X_{it}. As in conventional cross-sectional analysis, the problems stemming from ignoring omitted variables are markedly worse when the individual effects are correlated with the observed explanatory variables.[37]

33. For a concise derivation of the general result, see Amemiya 1985, 145–46. For introductory material and applications to panel data issues, see Greene 1997, 643–44; Baltagi 1995, 68–72; and Hsiao 1986, 48–49.

34. See Johnston and DiNardo 1997, 390.

35. See Stimson 1985.

36. For further details of this F-test, see Greene 1997, 617–18. For a discussion of the mean square error criteria for pooling, see Hsiao 1981, chap. 2; and Baltagi 1995, 54–60.

37. For further details, see Greene 1997; Johnston and DiNardo 1997; Hsiao 1986; and Baltagi 1995. We present a brief statistical introduction here.

The fixed-effects estimator is also known as the least-squares dummy variables (LSDV) estimator and the *within* estimator. As the name suggests, only the variation *within* each individual's observations is used. Since the number of dummy regressors will increase at the same rate as the number of individuals, N, it becomes impossible to estimate the resulting specification for all the unknown parameters. But by using the Frisch-Waugh-Lovell Theorem (FWLT), the estimates of the nondummy parameters can be obtained by conditioning Equation (6) on the individual effect dummies. This amounts to regressing the dependent variable and the nondummy regressors on the dummies and in turn running OLS on the resulting residuals to estimate the parameters of interest. The application of FWLT to Equation (6) is algebraically equivalent to deviating all variables from their individual means. The within-dyad means for the dependent variable and the vector of explanatory variables would be

$$\bar{y}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$$

and $\bar{X}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} X_{it}$, respectively,

with the fixed effects estimator taking the following form:

$$\hat{\beta}^{\mathrm{fe}} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \bar{X}_{i\cdot})(X_{it} - \bar{X}_{i\cdot})'\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \bar{X}_{i\cdot})(y_{it} - \bar{y}_{i\cdot})'\right]. \tag{7}$$

Rewriting Equation (6) to better represent the random-effects specification,

$$y_{it} = X_{it}\beta + \epsilon_{it} \text{ and } \epsilon_{it} = \alpha_i + u_{it}. \tag{8}$$

The disturbance term, ϵ_{it}, is composed of an individual-specific random component and an observation-specific random component. The presence of individual-specific random effects produces ϵ_{it}'s that are correlated across observations from the same individual. The generalized least squares (GLS) estimator is efficient under these conditions. In practice, a feasible generalized least squares (FGLS) estimator is used, since the disturbance variances are seldom known beforehand and require estimation. It can be shown that the random effects FGLS estimator is a weighted sum of the within estimator and the *between* estimator. The between estimator, as the name implies, uses the variation across individual means.

Note that if ϵ_i and X_{it} do not covary in Equations (6) and (8), omitting the individual effects from the estimated model (or presuming that they are homogeneous) would not produce biased coefficient estimates. In other words, under the null hypothesis of no correlation between individual effects and regressors, all three estimators—pooled OLS, fixed effects, and random effects—give consistent coefficient estimates. In the presence of individual effects uncorrelated with the regressors, the random-effects estimator would be consistent and efficient. If, however, the null is false and individual effects are correlated with the regressors, both the pooled OLS and random-effects estimators suffer from inconsistency. This provides the ingredients for a Hausman test for choosing between fixed and random effects. The test statistic becomes

$$\chi^2[k] = (\hat{\beta}^{fe} - \hat{\beta}^{re})'[Var(\hat{\beta}^{fe}) - Var(\hat{\beta}^{re})]^{-1}(\hat{\beta}^{fe} - \hat{\beta}^{re}), \quad (9)$$

where k is the number of time-variant regressors. If the fixed-effects specification were correct, as a large Hausman statistic would suggest, the fixed-effects estimator would, of course, be consistent. But also note that even if the random-effects specification were correct, the fixed-effects estimator would remain consistent—albeit inefficient—whereas the converse is not true.

Fixed-effects Logit (Conditional Logit)

Whereas much of the intuition from the linear panel model remains appropriate for models with binary dependent variables, the mechanics of estimation are different. The fixed-effects logit estimator used in the militarized-disputes model involves maximizing the *conditional* likelihood function,

$$L^c = \prod_{i=1}^{N} \text{Prob}\left(y_{i1}, \ldots, y_{iT} \middle| \sum_{t=1}^{T} y_{it} \right), \quad (10)$$

where $y_{it} = 0$ or 1. The joint probabilities for the individuals are conditioned on

$$\sum_{t=1}^{T} y_{it}$$

in order to sweep out the fixed effect, α_i.[38]

Under the null of homogeneity, the *unconditional* maximum likelihood estimator (MLE) is consistent and efficient. Under the alternate of heterogeneity, the conditional maximum likelihood estimator (CMLE) is consistent, whereas the MLE is inconsistent. Hence a Hausman test for heterogeneity can be performed using the following statistic:

38. For an accessible derivation of this estimator, see Baltagi 1995, 178–80. For a discussion of why observations from individuals that exhibit no variation over time in outcome do not contribute to the analysis (such as dyads that never experience militarized disputes), see Greene 1997, 899.

$$\chi^2[k] = (\hat{\beta}^{\mathrm{cmle}} - \hat{\beta}^{\mathrm{mle}})'[Var(\hat{\beta}^{\mathrm{cmle}}) - Var(\hat{\beta}^{\mathrm{mle}})]^{-1}(\hat{\beta}^{\mathrm{cmle}} - \hat{\beta}^{\mathrm{mle}}). \tag{11}$$

Likewise, we use a Hausman statistic constructed from the fixed-effects and random-effects logit estimators to test whether the individual effects are correlated with the regressors. This is identical to the test used in the linear trade model. Our estimation is based on the assumption that the individual effects are normally distributed.[39]

And Beyond . . .

The approaches outlined in this article do not exhaust the possible ways in which panel data can be studied. They are, however, the simplest to implement and hence the most commonly used and best understood approaches to empirical work. And since our main aim in this article is to encourage international relations scholars to make fuller use of the information embedded in the structure of their data, our fixed-effects analysis should be regarded as a first step toward developing richer models.

We have included discussion of two straightforward extensions of the fixed-effects model: the addition of period or year dummies and the modeling of dynamics. Since panel data possess both time and space dimensions, it is a logically symmetric extension to include a set of year dummies to our formulation. Time-series dynamics can be explicitly modeled as well.[40]

Improvements on the efficiency and inability of the fixed-effects estimator to analyze time-invariant variables are available under certain conditions. Hausman and Taylor develop a set of instrumental variables estimators for linear models that allow for the analysis of time-invariant regressors.[41] Although seldom used in applied work, their approach may be relevant to international relations research; researchers will have to investigate whether the requisite orthogonality conditions between subsets of regressors and fixed effects can be met with the variables available to them. In an attempt to improve on the efficiency of fixed-effects estimators, Gary Chamberlain proposes an alternate fixed-effects approach involving more precise specifications of the restrictions required when correlation is present between effects and regressors.[42]

Measurement error can pose a special problem in linear fixed-effects models, exacerbating the attenuation of estimated coefficients. However, Zvi Griliches and

39. For an introduction to random-effects logit models, see Hsiao 1986, 164–67; and Baltagi 1995, 178–82. For additional background material, see Pendergast et al. 1996; and Montgomery, Richards, and Braun 1986.

40. For a careful development of the two-way effects model, see Baltagi 1995, chap. 3. For an introduction to dynamic panel data analysis, see Greene 1997. For a discussion of more recent developments, see Baltagi 1995, chap. 8.

41. Hausman and Taylor 1981.

42. For a summary of Hausman and Taylor's arguments, see Hsiao 1986, 50–52; and Chamberlain 1984. For an introduction, see Johnston and DiNardo 1997, 404–408.

Hausman emphasize that it is not true in general that fixed-effects estimators are more biased than pooled estimators in the presence of measurement error.[43] In fact, if the variance of the measurement error is primarily cross-sectional, fixed-effects estimators may reduce the attenuation bias.[44] Using an instrumental variables approach applied to the trade model, we were unable to detect significant bias due to measurement error.

References

Amemiya, Takeshi. 1985. *Advanced Theory of Econometrics*. Cambridge, Mass.: Harvard University Press.

Anderson, James E. 1979. A Theoretical Foundation for the Gravity Equation. *American Economic Review* 69 (1):106–16.

Apodaca, Clair, and Michael Stohl. 1999. United States Human Rights Policy and Foreign Assistance. *International Studies Quarterly* 43 (1):185–98.

Auvinen, Juha. 1997. Political Conflict in Less Developed Countries 1981–1989. *Journal of Peace Research* 34 (2):177–95.

Baltagi, Badi H. 1995. *Econometric Analysis of Panel Data*. Chichester: John Wiley and Sons.

Barbieri, Katherine. 1996. Economic Interdependence: A Path to Peace or a Source of Interstate Conflict? *Journal of Peace Research* 33 (1):29–49.

Beck, Nathaniel, and Simon Jackman. 1998. Beyond Linearity by Default: Generalized Additive Models. *American Journal of Political Science* 42 (2):596–627.

Beck, Nathaniel, and Jonathan N. Katz. 1995. What to Do (and Not to Do) with Time-Series-Cross-Section Data. *American Political Science Review* 89 (3):634–47.

Beck, Nathaniel, and Richard Tucker. 1997. Conflict in Time and Space. Weatherhead Center for International Affairs Working Paper No. 97-8. Cambridge, Mass.: Harvard University. Available at <http://wwwc.cc.columbia.edu/sec/dlc/ciao/wps/tur01/>. Accessed July 1999.

Beck, Nathaniel, Jonathan Katz, and Richard Tucker. 1998. Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable. *American Journal of Political Science* 42 (4):1260–88.

Bennett, D. Scott. 1996. Security, Bargaining, and the End of Interstate Rivalry. *International Studies Quarterly* 40 (2):157–84.

———. 1997. Democracy, Regime Change, and Rivalry Termination. *International Interactions* 22 (4):369–97.

Benson, Michelle, and Jacek Kugler. 1998. Power Parity, Democracy, and the Severity of Internal Violence. *Journal of Conflict Resolution* 42 (2):196–209.

Bergstrand, Jeffrey H. 1985. The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence. *Review of Economics and Statistics* 67 (3):474–81.

Bernhard, William, and David Leblang. 1999. Democratic Institutions and Exchange-Rate Commitments. *International Organization* 53 (1):71–97.

Blanton, Shannon Lindsey. 1999. Instruments of Security or Tools of Repression? Arms Imports and Human Rights Conditions in Developing Countries. *Journal of Peace Research* 36 (2):233–44.

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

43. Griliches and Hausman 1986.
44. For an introduction, see Johnston and DiNardo 1997, 399–401.

Bremer, Stuart. 1996. Militarized Interstate Disputes, 1816–1992. Version 2.1. Available at ⟨http://pss.la.psu.edu/MID_DATA.HTM⟩. Accessed June 1997.

Chamberlain, Gary. 1984. Panel Data. In *Handbook of Econometrics,* edited by Zvi Griliches and Michael Intriligator, 1247–318. Amsterdam: North-Holland.

Clark, William Roberts, and Usha Nair Reichert, with Sandra Lynn Lomas and Kevin L. Parker. 1998. International and Domestic Constraints on Political Business Cycles in OECD Countries. *International Organization* 52 (1):87–120.

Dassel, Kurt, and Eric Reinhardt. 1999. Domestic Strife and Initiation of Violence at Home and Abroad. *American Journal of Political Science* 43 (1):56–85.

Deardorff, Alan V. 1984. Testing Trade Theories and Predicting Trade Flows. In *Handbook of International Economics*, edited by Ronald W. Jones and Peter B. Kenen, 467–517. Amsterdam: North-Holland.

Diehl, Paul F., Jennifer Reifschneider, and Paul R. Hensel. 1996. United Nations Intervention and Recurring Conflict. *International Organization* 50 (4):683–700.

Enterline, Andrew J. 1998. Regime Changes and Interstate Conflict, 1816–1992. *Political Research Quarterly* 51 (2):385–409.

Farber, Henry S., and Joanne Gowa. 1997. Common Interests or Common Polities? Reinterpreting the Democratic Peace. *The Journal of Politics* 59 (2):393–417.

Gartzke, Erik. 1998. Kant We All Just Get Along? Opportunity, Willingness, and the Origins of the Democratic Peace. *American Journal of Political Science* 42 (1):1–27.

Gowa, Joanne, and Edward D. Mansfield. 1993. Power Politics and International Trade. *American Political Science Review* 87 (2):408–20.

Greene, William H. 1997. *Econometric Analysis*. 3d ed. Upper Saddle River, N.J.: Prentice Hall.

Griliches, Zvi, and Jerry A. Hausman. 1986. Errors in Variables in Panel Data. *Journal of Econometrics* 31 (1):93–118.

Hall, Peter A., and Robert J. Franzese, Jr. 1998. Mixed Signals: Central Bank Independence, Coordinated Wage Bargaining, and European Monetary Union. *International Organization* 52 (3):505–35.

Hausman, J. A. 1978. Specification Tests in Econometrics. *Econometrica* 46 (6):1251–71.

Hausman, Jerry A., and W. E. Taylor. 1981. Panel Data and Unobservable Individual Effects. *Econometrica* 49 (6):1377–98.

Henderson, Errol Anthony. 1997. Culture or Contiguity: Ethnic Conflict, the Similarity of States, and the Onset of War, 1820–1989. *Journal of Conflict Resolution* 41 (5):649–68.

———. 1998. The Democratic Peace Through the Lens of Culture, 1820–1989. *International Studies Quarterly* 42 (3):461–84.

Heston, Alan, and Robert Summers. 1991. The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–88. *Quarterly Journal of Economics* 106 (2):327–68.

Hodgson, Geoffrey. 1996. An Evolutionary Theory of Long-Term Economic Growth. *International Studies Quarterly* 40 (3):391–410.

Hsiao, Cheng. 1986. *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Huber, Evelyne, Charles Ragin, and John D. Stephens. 1993. Social-Democracy, Christian Democracy, Constitutional Structure, and the Welfare-State. *American Journal of Sociology* 99 (3):711–49.

Huth, Paul K. 1996. Enduring Rivalries and Territorial Disputes, 1950–1990. *Conflict Management and Peace Science* 15 (1):7–41.

Iversen, Torben. 1998. Wage Bargaining, Central Bank Independence, and the Real Effects of Money. *International Organization* 52 (3):469–504.

Jackman, Simon. 1999. In and Out of War and Peace: The Statistical Analysis of Discrete Serial Data on International Conflict. Unpublished manuscript, Stanford University, Stanford, Calif.

Jaggers, Keith, and Ted Robert Gurr. 1995. Tracking Democracy's Third Wave with the Polity III Data. *Journal of Peace Research* 32 (4):469–82.

Johnston, Jack, and John DiNardo. 1997. *Econometric Methods*. 4th ed. New York: McGraw-Hill.

Jones, Daniel M., Stuart A. Bremer, and J. David Singer. 1996. Militarized Interstate Disputes, 1816–1992: Rationale, Coding, and Empirical Patterns. *Conflict Management and Peace Science* 15 (2):163–213.

Keith, Linda Camp. 1999. The United Nations International Covenant on Civil and Political Rights: Does It Make a Difference in Human Rights Behavior? *Journal of Peace Research* 36 (1):95–118.

Kim, Soo Yeon. 1998. Ties that Bind: The Role of Trade in International Conflict Processes, 1950–1992. Ph.D. diss., Yale University, New Haven, Conn.

King, Gary, and Langche Zeng. Forthcoming. Logistic Regression in Rare Events Data. *International Organization* 55 (4).

Leamer, Edward, and Herman Leonard. 1983. Reporting the Fragility of Regression Estimates. *Review of Economics and Statistics* 65 (2):306–17.

Leamer, Edward E., and Robert M. Stern. 1970. *Quantitative International Economics*. Boston: Allyn and Bacon.

Leblang, David A. 1996. Property, Democracy, and Economic Growth. *Political Research Quarterly* 49 (1):5–26.

———. 1997. Domestic and Systemic Determinants of Capital Control in the Developed and Developing World. *International Studies Quarterly* 41 (3):435–54.

Leeds, Brett Ashley, and David R. Davis. 1997. Domestic Political Vulnerability and International Disputes. *Journal of Conflict Resolution* 41 (6):814–34.

———. 1999. Beneath the Surface: Regime Type and International Interaction, 1953–1978. *Journal of Peace Research* 36 (1):5–21.

Lemke, Douglas, and William Reed. 1998. Power Is Not Satisfaction: A Comment on De Soysa, Oneal, and Park. *Journal of Conflict Resolution* 42 (4):511–16.

Lemke, Douglas, and Suzanne Werner. 1996. Power Parity, Commitment to Change, and War. *International Studies Quarterly* 40 (2):235–60.

Lieberson, Stanley. 1985. *Making It Count: The Improvement of Social Research and Theory*. Berkeley: University of California Press.

Linnemann, Hans. 1966. *An Econometric Study of International Trade Flows*. Amsterdam: North-Holland.

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., and Jack Snyder. 1997. A Tale of Two Democratic Peace Critiques: A Reply to Thompson and Tucker. *Journal of Conflict Resolution* 41 (3):457–61.

Meernik, James, Eric L. Krueger, and Steven C. Poe. 1998. Testing Models of U.S. Foreign Policy: Foreign Aid During and After the Cold War. *The Journal of Politics* 60 (1):63–85.

Montgomery, Mark R., Toni Richards, and Henry I. Braun. 1986. Child Health, Breast-Feeding, and Survival in Malaysia: A Random-Effects Logit Approach. *Journal of the American Statistical Association* 81 (394):297–309.

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):649–61.

Mousseau, Michael. 1997. Democracy and Militarized Interstate Collaboration. *Journal of Peace Research* 34 (1):73–87.

Noland, Marcus. 1997. Chasing Phantoms: The Political Economy of USTR. *International Organization* 51 (3):365–87.

Oneal, John R., and James Lee Ray. 1997. New Tests of the Democratic Peace: Controlling for Economic Interdependence, 1950–1985. *Political Research Quarterly* 50 (4):751–75.

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: Interdependence, Democracy, and International Conflict, 1950–85. *Journal of Peace Research* 33 (1):11–28.

Pendergast, Jane F., Stephen J. Gange, Michael A. Newton, Mary J. Lindstrom, Mari Palta, and Marian R. Fisher. 1996. A Survey of Methods for Analyzing Clustered Binary Response Data. *International Statistical Review* 64 (1):89–118.

Perry, Robert L., and John D. Robertson. 1998. Political Markets, Bond Markets, and the Effects of Uncertainty: A Cross-national Analysis. *International Studies Quarterly* 42 (1):131–59.

Poe, Steven C., C. Neal Tate, and Linda Camp Keith. 1999. Repression of the Human Right to Personal Integrity Revisited: A Global Cross-National Study Covering the Years 1976–1993. *International Studies Quarterly* 43 (2):291–313.

Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies.* Berkeley: University of California Press.

Raknerud, Arvid, and Håvard Hegre. 1997. The Hazard of War: Reassessing Evidence of the Democratic Peace. *Journal of Peace Research* 34 (4):385–404.

Remmer, Karen L. 1998. Does Democracy Promote Interstate Cooperation? Lessons from the Mercosur Region. *International Studies Quarterly* 42 (1):25–42.

Russett, Bruce. 1978. The Marginal Utility of Income Transfers to the Third World. *International Organization* 32 (4):913–28.

Russett, Bruce, John R. Oneal, and David R. Davis. 1998. The Third Leg of the Kantian Tripod for Peace: International Organizations and Militarized Disputes, 1950–1985. *International Organization* 52 (3):441–67.

Signorino, Curtis A. 1999. Strategic Interaction and the Statistical Analysis of International Conflict. *American Political Science Review* 93 (2):279–97.

Simmons, Beth A. 1996. Rulers of the Game: Central Bank Independence During the Interwar Years. *International Organization* 50 (3):407–43.

Singer, J. David. 1990. *Models, Methods, and Progress in World Politics: A Peace Research Odyssey.* Boulder, Colo.: Westview Press.

Stimson, James A. 1985. Regression in Space and Time: A Statistical Essay. *American Journal of Political Science* 29 (4):914–47.

Thompson, William R., and Richard Tucker. 1997a. A Tale of Two Democratic Peace Critiques. *Journal of Conflict Resolution* 41 (3):428–54.

———. 1997b. Bewitched, Bothered, and Bewildered: A Reply to Farber and Gowa and to Mansfield and Snyder. *Journal of Conflict Resolution* 41 (3):462–77.

Tinbergen, Jan. 1962. *Shaping the World Economy.* New York: Twentieth Century Fund.

Tir, Jaroslav, and Paul F. Diehl. 1998. Demographic Pressure and Interstate Conflict: Linking Population Growth and Density to Militarized Disputes and Wars, 1930–89. *Journal of Peace Research* 35 (3):319–39.

Wang, T. Y. 1999. U.S. Foreign Aid and UN Voting: An Analysis of Important Issues. *International Studies Quarterly* 43 (1):199–210.

Ward, Michael D., and Kristian S. Gleditsch. 1998. Democratizing for Peace. *American Political Science Review* 92 (1):51–61.

Zahariadis, Nikolaos. 1997. Why State Subsidies? Evidence from European Community Countries, 1981–1986. *International Studies Quarterly* 41 (2):341–54.