Tradeoffs in Trade Data

*Do Our Assumptions Affect Our Results?**

CHARLES R. BOEHMER

University of Texas at El Paso

BERNADETTE M. E. JUNGBLUT

West Virginia University

RICHARD J. STOLL

Rice University

Researchers investigating the link between trade and peace often face a severe problem of list-wise deletion from missing trade data. Attempts to mitigate this problem include assuming that most observations are zero or imputing the values of such flows. We compare two frequently used trade data sets (the Gleditsch data set and the Correlates of War Project data set). We classify individual observations as observed, constructed or missing. We demonstrate that state attributes are systematically related to different categories of trade data. Using Monte Carlo simulations, we also find that replacing some missing data with estimated values tends to inflate the effects of trade in conflict models, although the effects differ by data set.

KEYWORDS: data assumptions; economic data; international conflict; international trade; missing data; monte carlo simulations; trade data

Introduction

Research on trade interdependence and interstate conflict has become an important and active area of study. The findings to date, however, have been mixed. Divergent findings have been tied to differences in empirical estimation, variable construction, model specifications, and differences in trade data sets. Our study investigates how missing trade data and the methods to replace missing observations may influence the empirical findings that are at the heart of these debates. If trade data are missing completely at random, then this would be an ignorable problem for our estimates

*We would like to thank Brian Pollins, John Oneal, Katherine Barbieri, Rafael Rueveny, Doug Block, and three anonymous reviewers for help and suggestions with this project. Earlier versions of this paper were presented at the meetings of the Peace Science Society in 2005 (Iowa) and 2008 (Claremont).

as long as we still have more observations than covariates.¹ If, however, trade data, or data for any other independent variable, are missing systematically, then this can reduce statistical variance, lead to statistical insignificance in hypothesis testing and even bias estimates (Allison, 2002). Conflict scholars have found it necessary to reduce missing data by constructing trade data using a variety of techniques and decision-rules. But decisions to construct trade values are based on assumptions about what trade values would look like if we had excellent documentation combined with consistent collection techniques and coding rules. Examining the two most widely used trade data sets in international relations (Gleditsch, 2004, and Barbieri, Keshk, and Pollins, 2008; hereafter COW/BKP), we find that the assumptions one makes to replace missing data indeed appear to affect estimates of the effects of bilateral trade flows on interstate conflict.²

Debates over Trade Model Specification, Estimation, and Data

Studies examining the effects of trade on interstate conflict have produced divergent findings. Barbieri (1995, 1996, 2002) has argued that trade increases interstate conflict, while many other studies find supporting evidence for the “liberal peace” assertion that trade reduces the probability of such conflict (Oneal et al., 1996; Oneal and Russett, 1997, 1999a,b; Oneal et al., 2003; Russett and Oneal, 2001). Oneal (2003: 724) suggests this debate is settled given the 30 or so studies by 25 different scholars providing evidence showing the pacifying effects of trade. Instead, debate has turned to matters of model specification, estimation, and data. Special attention has been paid to how to treat missing data and variable construction.

Some scholars have suggested these divergent findings are a function of potentially improper statistical analysis techniques. Beck et al. (1998) find that when using their “peace years” and splines technique, the effect of trade on conflict is statistically insignificant. Debate has also included the appropriateness of fixed effects estimators (Green et al., 2001; Oneal and Russett, 2001; Beck and Katz, 2001; King, 2001). Keshk et al. (2004, 2010) use simultaneous equations to examine the endogenous relationship between interstate trade and militarized conflict. They find evidence that conflict reduces trade but inconsistent evidence that trade flows in turn reduce militarized conflicts.

Other researchers speak to this debate by examining the different measures of trade used by Barbieri and Oneal and Russett. Gartzke and Li (2003a) distinguish among “trade share”, “trade dependence”, and “trade openness” and investigate their effects on militarized interstate dispute (MID) onset. Examining the 1950–1992 time period, Gartzke and Li construct these measures from both the trade data sets of Barbieri (1998) and Oneal and Russett (1999a). Gartzke and Li’s findings,

¹ This is the concept of “micronumerosity”. See Gujarati (2003: ch. 10) for a discussion.

² One reviewer expressed concern about the possibility of non-stationarity of trade data; however, investigating that issue and its impact on the findings to date regarding the trade–conflict relationship would entail a large, separate project, and is not the focus of the research presented here.

however, have not settled the substantive debate. Both Oneal (2003) and Barbieri and Peters (2003) take issue with Gartzke and Li's conclusions. Gartzke and Li (2003b) find Oneal's and Barbieri and Peters's critiques unpersuasive. Hegre (2005a) also takes issue with Gartzke and Li's (2003a,b) interpretation of their findings, focusing specifically on their assertion that while "trade share" measures a state's disconnectedness from world trade, "trade dependence" captures a state's connectedness to world trade. Hegre (2005a) is also concerned that Gartzke and Li's trade openness measure may be masking the effects of the size of states' economies. This debate remains unresolved, as Hegre (2005b) does not find Gartzke and Li's (2005) response to his critiques sufficient or compelling.

Most relevant for our purposes is how the problem of missing data has been treated in past studies. Oneal and Russett code missing dyadic trade data from the IMF as zero, although neither the original Barbieri data set nor the Barbieri, Keshk, and Pollins³ data set (COW/BKP) follow this decision-rule. In an attempt to determine if these differences in the treatment of missing trade data have a substantive effect, Gartzke and Li construct two versions of Oneal and Russett's data—the first with missing trade values coded as zero and the second with missing trade values coded as missing. Across these different data sets, Gartzke and Li find that while trade dependence and trade openness reduce the likelihood of MID onset, conversely, trade share increases the likelihood of MID onset. Gelpi and Grieco (2008) show that the effect of trade dependence, using trade data from both Oneal and Russett and the original Barbieri data set, appears contingent on the level of democracy of the potential initiator of interstate conflict and the trade dependence of the potential target.

Of primary importance for our purposes is Gelpi and Grieco's imputation of missing values following King et al. (2001). Their results also lead them to suggest that Barbieri's original finding that trade dependence increases militarized interstate disputes was based on a sample bias stemming from missing data. They also believe their imputation method is more accurate than treating all missing values as zero, as do Oneal and Russett in much of their work. Keshk et al. (2010) examine the robustness of different treatments for missing data in their study by dropping observations where IMF data equaled zero or replacing them with a nominal value. The results were congruent with their 2004 study that international relations appears to drive trade flows and not the reverse. Another study is of additional note here concerning research design. Gartzke (2007) argues that the liberal peace should be interpreted as a "capitalist peace". He provides evidence that peace is a function of financial openness and level of development, washing out the effects of democracy and trade. However, his incorporation of three economic variables (trade flows, financial data, and GDP) appears to compound the problem of list-wise deletion. It is thus not unreasonable to suspect that his results are at least in part a function of sample bias if missing data are missing systematically (Dafoe, 2008).

The findings of all the studies above are affected not only by important assumptions regarding variable construction but also by the quality of data

³ The BKP data are available through the Correlates of War Project. Others refer to these data as the Correlates of War Project trade data.

employed. Are we drawing accurate and reliable inferences in regard to both the direction and substantive effects of trade on conflict? Our primary goal here is to analyze the effects of missing data and their implications on the trade–conflict relationship, suspecting that missing data are masking our ability to discern patterns of theoretical interest. This is an important step towards understanding the reasons for the discrepant findings discussed above.

Categorizing Types of Trade Data

Anyone that has collected or scrutinized trade data is keenly aware of the thorny problems relating to sources and missing data. Those that have collected and refined trade data should be applauded for their efforts, but in doing so they have constructed trade data to replace missing trade figures. We categorize three types of trade data in the Gleditsch and Correlates of War (BKP) trade data sets. There are *observed* trade figures. These figures are taken from various official sources. There are *constructed* trade figures. These figures are generated based on various assumptions, whether explicit or implicit, by Gleditsch and COW/BKP. To be clear, we classify as constructed any trade figure that is the result of using any procedure to replace a missing trade data observation.⁴ Finally, some data are *missing*.⁵ Table 1 displays the three-way categorization of trade data for the Gleditsch and COW/BKP trade data.

This leads us to ask three questions about these trade figures. First, do the characteristics of states vary with the type of trade data? That is, are states with observed trade data systematically different from states with constructed trade data and/or from states with missing trade data? If so, then decisions about including and excluding cases based on the type of trade data may have an effect on one's analyses and results. Gleditsch himself notes that trade data are not missing at random; for example he says "...bilateral trade flow data tend to be poor—if at all available—for smaller, developing countries that are far apart or pairs of antagonistic states" (Gleditsch, 2002: 712). Barbieri (2002) agrees there appears to be a systematic pattern to the missing data. Not that this is not just a potential problem for trade data but a problem for all economic variables studied using time-series cross-sectional (TSCS) research designs. Naturally, the farther one goes back in time the more one should expect economic data to be missing for states.

Second, are missing exports and imports data distributed similarly? One would think that missing exports and imports data in bilateral observations would square

⁴ We adopt a very strict definition of 'constructed'. Anytime a researcher replaces a missing trade data figure, we call this constructed. This includes the following: if A's imports from B are missing, but there is a recorded figure for B's exports to A and that figure is used to replace the missing import figure, we call this constructed. Other researchers may feel that since the missing data is being replaced by a recorded figure, this is not constructed data. We acknowledge this to be a reasonable position, but we prefer to use our strict definition.

⁵ The Gleditsch data observation codes are listed in the file `readme_4.1.asc`, which is part of the .zip file that can be downloaded at <http://weber.ucsd.edu/~kgledits/exptradegdp.html>. The missing data codes for the COW/BKP data set are available in their Correlates of War Project Trade Data Set Codebook, Version 2.01 (<http://correlatesofwar.org>).

Table 1. Gleditsch and COW/BKP: Categories of Trade Data

<i>Gleditsch: Exports</i>	<i>Gleditsch: Imports</i>	<i>Category</i>	<i>BKP</i>	<i>Category</i>
0 Observed data from IMF Direction of Trade (DOT) Data	0 Observed data from IMF Direction of Trade (DOT) Data	Observed	1. Barbieri version 1	Observed
1 Observed data from World Export Database (WED)	1 Observed data from World Export Database (WED)	Observed	2. IMF imports, c.i.f.	Observed
1.1 Observed data from the Statistical Yearbook of ROC (Taiwan)	1.1 Observed data from the Statistical Yearbook of ROC (Taiwan)	Observed	3. Missing replaced with exporter, f.o.b.	Constructed
1.2 Observed data from Directorate General of Budget Accounting and Statistics Executive Yuan of the ROC (Taiwan)	1.2 Observed data from Directorate General of Budget Accounting and Statistics Executive Yuan of the ROC (Taiwan)	Observed	4. Zero replaced with exporter, f.o.b.	Constructed
2 Estimate based on B's imports from A (observed data from DOT)	2 Estimate based on B's exports to A (observed data from DOT)	Constructed	5. Missing replaced with importer, c.i.f.	Constructed
3 Estimate based on A's imports from B	3 Estimate based on A's exports to B	Constructed	6. Zero replaced with importer, c.i.f.	Constructed
4 Interpolated estimate	4 Interpolated estimate	Constructed	7. Missing replaced with exporter, f.o.b.	Constructed
5 Lags and leads based on first non-missing observations, deflated to current prices	5 Lags and leads based on first non-missing observations, deflated to current prices	Constructed	8. Zero replaced with exporter, f.o.b.	Constructed
6 Dyads with no observed data, assumed to be 0	6 Dyads with no observed data, assumed to be 0	Constructed	9. Missing replaced with Barbieri version 1	Constructed
			10. Zero replaced with Barbieri version 1	Constructed
			11. Belgium-Luxembourg BEL-LUX data	Observed
			12. Taiwan from multiple sources	Observed
			13. Aggregate People's Republic of China (CEN), Hong Kong (HK), Macao	Observed
			14. Missing	Missing

There are no missing data in the Gleditsch trade data.

with each other; however, this is rarely the case. Instead, the values of imports and exports often do not match, or some observations are missing. There may be many reasons for this problem, ranging from mundane issues with historical records and references to the more interesting and politically motivated intentional misreporting strategies. See Barbieri et al. (2009) for a detailed discussion of this issue. We do not offer any thorough explanation of this problem; instead we explore how significant a problem this is overall.

Third, if we replace missing trade data with constructed trade data, does this alter the results of trade–conflict analyses (or any other relationship for that matter), and if so, in what ways? Gleditsch also provides some insight into this question in his article. He studies the influence of dyadic economic dependence and distance on the likelihood of a militarized interstate dispute, comparing results using only observed trade data with results using observed and estimated trade data. He finds that the substantive effect of trade was higher in the more complete sample (with missing data replaced) relative to the observed sample, in part related to conflicts that were systematically missing such as those between countries in the Eastern Bloc. We later use Monte Carlo simulations of both the COW/BKP and Gleditsch data sets to demonstrate how our model estimates of militarized conflicts vary as constructed trade data replace missing trade data.

The Gleditsch Trade Data

The Gleditsch trade data covers the years 1948–2000. The trade data are scaled in millions of current US dollars, and the unit of analysis is the dyad-year. Each record contains four trade figures for the dyad A–B: exports of State A to State B, imports of State A from State B, exports of State B to State A, and imports of State B from State A. The bulk of the observed trade data originates from the *IMF Direction of Trade Data*. A smaller amount of observed trade is taken from other official sources. The manner in which Gleditsch estimates trade data is described in both his published article (Gleditsch, 2002: 716–718) and in the documentation file for the data set. For our purposes, we consider observed trade data to be those data Gleditsch obtained from the *IMF Direction of Trade Data* and his other cited sources (i.e. data categories 0, 1, 1.1, and 1.2). Constructed data are those data he designates as “estimated” in his documentation and the data set itself. The sources for Gleditsch’s trade data (and their categorization as observed or constructed) are given in Table 1.

As can be seen in Table 1, Gleditsch uses a variety of techniques to construct trade data. If an import figure is missing, but the corresponding export figure is available (i.e. if A’s imports from B is missing, but B’s exports to A is available), the export figure replaces the missing import figure and vice versa (substitute available import data for missing export data). He uses interpolation to replace missing trade data. He also uses lags and leads to cover ranges of missing data. Finally, in a number of cases he codes a value of zero if his research indicated that it is unlikely there was any trade between states. There are *no* missing data in the Gleditsch trade data set. More information on his trade data is available in Gleditsch (2002), especially pp. 716–718.

The Barbieri, Keshk, and Pollins (COW/BKP) Trade Data

The Correlates of War Trade Data Set, version 2.01, covers the years 1870–2006 in both monadic and dyadic forms. Two trade flows are measured: the imports from state A to state B (flow1), and the imports from state B to state A (flow2). Trade flows are measured in current millions of US dollars (Barbieri et al., 2008). The most obvious difference between the COW/BKP data set and the Gleditsch data set is the temporal domain. But a far more significant difference is the treatment of missing data in the two data sets.

As we noted, Gleditsch replaces *all* missing trade data with constructed data. BKP are far more conservative in their approach to missing trade data, but they do create some constructed trade data. They often rely on using the imports from one state to replace exports in their trade partner, which they reason minimizes the chance of error in the data because states are more likely to misrepresent or misreport exports than imports. When only one figure is available for an import/export pair, the available figure is substituted for the missing data figure. Finally, missing data and zero values for IMF trade data were replaced by values from Barbieri's (1998) trade data set. But they do not replace missing trade data with zeros (see Barbieri et al., 2009, for a discussion of their reasoning). They also argue against interpolating or extrapolating trade data.

Due to the longer time period of the BKP data set, they have more observed trade data observations than Gleditsch. But they also have a significant amount of missing data. About 30 percent of their trade data are missing. We want to stress that the amount of missing data in their data set is a *deliberate decision* by BKP. They have a very different attitude than Gleditsch on what techniques are appropriate to create constructed trade data.

Predicting Categories of Trade Data by State Characteristics

Are states with observed trade data, constructed trade data, and missing trade data different? Can we associate each of our three categories of trade figures with state characteristics? The authors of the trade data sets both expect that, as would be commonly anticipated by others who have worked with time-series cross-sectional (TSCS) economic data, some data are missing systematically, although there has been little empirical investigation of this point. For example, Mansfield et al. (2000: 317) note that they encountered problems collecting trade data for some autocracies that bias their empirical results in a manner that “understates” their findings. This is consistent with the expectations by Allison (2002). However, few if any past studies have documented how state characteristics are related to all three of our trade categories (observed, constructed, and missing). If this is the case, then it is possible the results we obtain by using only observed trade data are not representative of what we would obtain if there were no missing trade data. A problem may occur using list-wise deletion if data are missing systematically, resulting in higher standard errors and thus inaccurate test statistics (Allison, 2002: 6). We begin to answer this question by seeking to characterize the differences between states that have observed trade data figures, states that have trade data figures constructed by Gleditsch or BKP, and states that have missing trade data figures. To maximize the comparability of the results, we analyze both trade data sets from 1948 on.

We selected a series of state characteristics we felt would provide a good prediction of missing trade figures. Some have a prominent place in analysis of the democracy–trade–conflict literature. Others have a prominent place in the study of comparative foreign policy. This is by no means an exhaustive list of potential variables that may affect the availability of trade data. Not included in the list below is a variable for participation in militarized interstate disputes or wars. We agree that some trade flows are disrupted by conflict or trade figures are not recorded during conflicts. Because interstate conflict is a function of other state and bilateral characteristics, we do not include it here and instead refer readers to Barbieri et al. (2009) for evidence to this effect. Because it should be acknowledged that data are more difficult to obtain in the distant past, “time” is another variable we do not explore in our analysis, although we expect that trade figures are more available going back in time for states which vary positively with the variables below. The state characteristics we selected below were extracted using EUGene (Bennett and Stam, 2000):⁶

- the level of democracy
- the size of the economy (operationalized as commercial energy consumption)
- the level of development (operationalized as commercial energy consumption per capita)
- the Correlates of War Project power capability score (i.e. CINC score)
- major power status
- region

The case structure for this analysis uses all non-directed dyads. Each of the first four state characteristics above serves as a separate dependent variable in a regression equation. Because each dependent variable is a continuous measure, we use OLS regression but omit the constant. We also show using contingency tables how observed, constructed, and missing data are distributed by major power status and region for each of the Gleditsch and BKP data sets. We present these latter results in the appendix. (See Appendix Tables A1, A2, and A3.) In each regression equation we include three binary independent variables, one for each of our categories of trade data: Observed, Constructed, and Missing. This results in 12 total trade dummy

⁶ The democracy score, *dem*, was generated by subtracting the Polity IV autocracy score from the Polity IV democracy score for a state. Note that in the generation of his trade data, Gleditsch used the list of states from Gleditsch and Ward (1999). This list contains more entities (and/or has entities with a longer period of years) than exist in the Correlates of War Project state list. Consequently, there are a number of state-years in the Gleditsch trade data for which we do not have data from EUGene. Our use of energy consumption as an indicator of the size of the economy (as opposed to using GDP data) and energy consumption per capita as an indicator of development (as opposed to GDP per capita) is consistent with a great deal of previous research in international relations. But there is another reason we choose to use energy consumption instead of GDP. Just like trade data, some GDP data are from official sources, some are constructed, and some are missing. Given that the purpose of this study is to explore the implications of varying the amount of constructed trade data (versus missing data), we felt that adding another variable with the same issues would only serve to complicate (and possibly obscure) the focus of our study.

Table 2. Regression Slope Coefficients for Continuous State Characteristics Gleditsch Data

	Export Data: State Characteristic				Import Data: State Characteristic			
	Dem	Energy	Energy Per Capita	CINC	Dem	Energy	Energy Per Capita	CINC
Observed	0.996	86418.8	2.913	0.010	1.545	86532.5	3.007	0.009
Constructed	-1.675	34934.8	2.539	0.004	-1.775	45352.9	2.496	0.007
Missing	-2.111	18632.6	2.094	0.002	-2.111	18632.6	2.094	0.002

Table 3. Regression Slope Coefficients for Continuous State Characteristics COW/BKP Data

	Flow 1: State Characteristic				Flow 2: State Characteristic			
	Dem	Energy	Energy Per Capita	CINC	Dem	Energy	Energy Per Capita	CINC
Observed	2.120	97972.0	2.849	0.009	1.646	106248	2.952	0.010
Constructed	-1.881	77133.0	2.580	0.010	1.300	60669	2.369	0.007
Missing	-1.630	31024.9	1.838	0.004	-1.744	24488	1.712	0.003

variables, three each for the Gleditsch export data, the Gleditsch import data, the BKP flow1 data, and the BKP flow2 data. Please see the appendix for the structure of the case records and how we convert the dyadic records to state-level records. We provide one model specification below from our analysis to illustrate what the models look like:

$$\text{Democracy} = \text{Imports_observed} + \text{Imports_constructed} + \text{Imports_missing} + e$$

Tables 2 (the Gleditsch data) and 3 (the COW/BKP data) summarize the results of the regression analysis of the continuous state characteristics. The full estimates of these models are available from the authors upon request. Given the large number of observations (over one million), we would expect all coefficients to be statistically significant. This indeed is the case so the tables only show the regression coefficients. The dependent variables for each model are listed in the columns and the slope coefficients for each trade dummy independent variable fill the cells.

In terms of the Gleditsch data, there is a consistent ordering by trade data. States with observed trade data are on average more democratic, have a higher level of energy consumption, a higher level of energy consumption per capita, and higher CINC scores. The slope coefficients for the observed Gleditsch import and export data are always higher than those for the constructed and missing categories. States with constructed data score lower on all these variables. States with missing trade data have the lowest values on all of these variables. Thus, these states are the least democratic, have the smallest economies, and are the least developed and least powerful. Also of interest is the sign change on the level of democracy variable as we move from observed data (positive) to estimated or

missing data (negative). Moving from observed data to estimated or missing data, there are large magnitude changes in the size of both the economy variable and the power capability variable.

In terms of the COW/BKP data, missing data are still associated with less democratic states, smaller economies, less developed states, and less powerful states. Again, we observe a sign change for democracy as we move from observed flow1 data to estimated or missing flow1 data—but only when we change from observed flow2 data to missing flow2 data. The sign change is not in effect for flow2 estimated data. As with the Gleditsch trade data, there are large magnitude changes in the size of the economy and power capability as we move from observed to missing or constructed data. Of particular interest is a large change in magnitude for the level of development when using the COW/BKP data. This is a factor that may likely affect the sample of states included and thus likely trade–conflict estimates.

Is this important? Possibly. Replacing missing trade data with constructed figures changes the mix of state characteristics in one's sample. This in turn may change the estimated relationships in one's analysis. The variance for trade variables in a sample that contains only observed trade data should be artificially low when missing trade figures are missing systematically due to their dependence on the values of other independent variables. Allison notes that such data are not missing at random (MAR), and under such conditions, regression estimates that treat missing data by list-wise deletion inflate standard errors. This could render slope coefficients statistically insignificant. Interestingly, list-wise deletion combined with logistic regression can still yield robust and accurate estimation as long as “the probability of missing data on any variable depends on the dependent variable but does *not* depend on any of the independent variables” (Allison, 2002: 7). Our analysis shows this is not the case; missing trade data can be predicted based on the values of other independent variables often accompanying a trade variable. Hence, estimates of interstate conflict may not be very reliable from sample to sample and with changes in model specifications due, in part, to list-wise deletion, which is magnified for dyadic studies because if either state in a dyad is missing trade data this observation will drop out of the estimation. Another implication is that dyadic studies will be missing proportionally more observations than state-level analyses if using only observed data. Again, much of this depends on the sample one begins with and then methods used to fill in missing data but also depends on both theory and trade measures.

However, both the BKP and Gleditsch data sets include constructed trade figures, and this makes it less clear how much the problem of list-wise deletion affects estimates of interstate conflict or other dependent variables. Adding back some constructed trade figures could improve the variance of trade variables that reduces non-random sampling error. In regard to the method used by Gleditsch, we offer no critique here but instead focus on understanding the data. BKP (2009), however, demonstrate how the balanced trade relationship assumption could yield artificial symmetry in the aggregated trade variables. Our analysis also shows that there are differences between exports and imports or flows that point to a lack of symmetry in trade data. Although incorporating constructed trade figures could improve the efficiency of estimation, there is little theory to guide us about what particular

method of construction should be used on which particular missing observations. If imputation and other methods are used, there is an increased risk of increasing error because we often do not have very good theoretical understanding about why data are missing. Our next section provides a test to see how the inclusion of constructed trade figures affects estimates of militarized interstate conflicts.

Constructed Trade Data and its Effects on the Estimation of Militarized Conflict

Here we seek to extend our understanding of the impact of the choices made in constructing trade data sets on our substantive understanding of the trade–conflict relationship. Specifically, how do differences among “observed”, “constructed”, and “missing” trade data affect the trade–conflict relationship? We now turn to an examination of the influence of using trade data estimates on the relationship between trade dependence and conflict. This is a rather important topic methodologically because it strikes at the heart of assumptions in our research designs. Allison (2002) says of imputation, “the basic idea is to substitute some reasonable guess for each missing value and then proceed to do the analysis as if there were no missing data” (p. 11). He goes on to describe several methods that may be used to impute data and under what conditions. Imputation is particularly more complicated when data are missing in more than one independent variable, which based on the results shown in the previous section would appear to be a risk for our models of trade and interstate conflict. In terms of the various methods available, Allison warns us that “unfortunately, all of these imputation methods suffer from a fundamental problem: Analyzing imputed data as though it were complete data produces standard errors that are underestimated and test statistics that are overestimated. Conventional analytic methods simply do not adjust for the fact that the imputation process involves uncertainty about the missing values” (Allison, 2002: 12).

Gleditsch demonstrates this effect in his own data in a simple analysis (2002: 720). Using the dyad-year design, he predicts the onset of militarized interstate disputes from a logit analysis using trade dependence and the natural log of distance between the capital cities of the dyad. His results are included in our appendix (see Appendix Table A4). He shows that the trade dependence is about 50% larger and that there are of course many more cases in the model with imputed data. We want to conduct a more thorough examination of the impact of replacing missing trade data with estimated trade data. We place the trade data in the context of a full model predicting conflict by varying the amount of estimated trade data so that we can track the impact of replacing missing data with trade data estimates. We base our Monte Carlo work on the analysis of Russett and Oneal (2001).⁷

We wanted to base our study on published research that found a significant relationship between trade dependence and conflict. There are a number of different pieces we could have used, and our choice of this particular set of results was somewhat arbitrary. On the other hand, Russett and Oneal’s body of work is

⁷ The Stata command used to replicate these results is: `xtgee dispute1 smlmat smldep smigo-
abi lcaprat allies noncontg logdstab minrpwrs, family(binomial) link(logit) corr(ar1) force
robust nolog.`

widely known and their book *Triangulating Peace* is widely cited.⁸ So most researchers interested in the impact of interdependence on conflict will be familiar with their results.

We begin by “replicating” this analysis for both sets of trade data.⁹ We use the Russett and Oneal data from their book, except that we alter the lower dependence variable. This variable is created in two steps. First, for each state in a dyad, the ratio of its total trade with the other state to the state’s GDP is calculated. Second, the *lower* of these two ratios is used in the analysis. In our analysis, we create two series of lower dependence variables, one using Gleditsch’s trade data and the other using the COW/BKP trade data. For our variables, we use Gleditsch’s GDP data for the denominator of our indexes.

We start by using only the *observed* trade data from each data set. These analyses have the maximum amount of missing trade data from each data set. The results of the analysis using each trade data set are shown in Table 4. We then “replicate” this analysis including all observed and *constructed* data from each trade data set. These analyses minimize the missing data and the results are also shown in Table 5.

In terms of the Gleditsch data results shown in Tables 4 and 5, our interest is in the coefficient for lower trade dependence. First, when only observed data are used, lower trade dependence is statistically insignificant. One reason for this may be a reduction in the variance of trade as the variable becomes more homogenous due to missing data. Restricted variance in a right-hand side variable is one reason for insignificant regression coefficients (Lewis-Beck, 1980: 35). We should expect this if trade data are more likely to be missing for some types of state compared to others, as we demonstrated above. In contrast, when both observed and constructed trade data are included, lower trade dependence is negative and statistically significant. Moreover, as we move from observed to observed plus constructed Gleditsch trade data, the size (absolute value) of the lower trade dependence coefficient doubles. In contrast, when we examine the COW/BKP data, we find that for both the observed and observed plus constructed trade data, the lower trade dependence coefficient is negative and statistically significant. The magnitude of change from observed trade data (–37.974) to observed plus constructed trade data (–60.197) is about 1.5.

We would like to provide a more explicit characterization of the impact of replacing missing data with constructed data. A Monte Carlo study varying the amount of constructed data is an appropriate way to assess the impact of varying amounts of constructed data on substantive findings. Our Monte Carlo study is conducted as follows. Beginning with only the observed trade data (i.e. the data

⁸ A search in Google Scholar turns up over 700 citations to the book (search conducted 24 September 2009).

⁹ It should be noted that in their analysis, Russett and Oneal do not use the Gleditsch or COW/BKP trade data. They begin with IMF trade data. They supplement these data with Barbieri’s (1998) trade data, replacing missing data with her figures when possible. Finally, because members of the IMF are required to report their trade if either state in a dyad is a member of the IMF, if there was no entry for their trade, Russett and Oneal substituted a value of zero (Russett and Oneal, 2001: 139).

Table 4. Russett and Oneal Table A5.1 Using Gleditsch and COW/BKP Trade Data: Observed Data Only

MID Onset	Gleditsch Observed Trade Data			COW-BKP Observed Data		
	Coef.	Std. Err.	$P > z $	Coef.	Std. Err.	$P > z $
Lower democracy	-0.051	0.015	0.000	-0.043	0.010	0.000
Lower dependence	-41.20	33.10	0.213	-37.974	16.601	0.022
Int'l organizations	0.001	0.006	0.814	-0.010	0.004	0.023
Log of power ratio	-0.080	0.072	0.269	-0.278	0.059	0.000
Allies	-0.462	0.281	0.100	-0.275	0.204	0.177
Non-contiguity	-1.330	0.289	0.000	-1.165	0.169	0.000
Log of distance	-0.265	0.092	0.004	-0.287	0.070	0.000
Only minor powers	0.269	0.299	0.369	-0.285	0.214	0.182
Constant	-1.941	0.783	0.013	-1.098	0.562	0.051
GEE population-averaged model						
Number of groups		790			1,000	
N		16,587			25,776	

Table 5. Russett and Oneal Table A5.1 Using Gleditsch and COW/BKP Trade Data: Observed and Constructed Data

MID Onset	Gleditsch Observed + Constructed			COW-BKP Observed + Constructed		
	Coef.	Std. Err.	$P > z $	Coef.	Std. Err.	$P > z $
Lower democracy	-0.046	0.013	0.001	-0.058	0.011	0.000
Lower dependence	-84.8	40.8	0.038	-60.2	20.562	0.000
Int'l organizations	-0.024	0.007	0.000	-0.011	0.005	0.023
Log of power ratio	-0.366	0.063	0.000	-0.327	0.056	0.000
Allies	-0.784	0.211	0.000	-0.536	0.188	0.004
Non-contiguity	-1.252	0.258	0.000	-1.104	0.197	0.000
Log of distance	-0.450	0.084	0.000	-0.325	0.077	0.000
Only minor powers	-0.874	0.274	0.001	-0.596	0.212	0.005
Constant	0.871	0.721	0.227	-0.618	0.652	0.343
GEE population-averaged model						
Number of groups		944			1,060	
N		28,177			33,809	

used to generate the results of Table 4), we randomly add increments of 5% of constructed data (i.e. we first replace 5% of the missing data with constructed data, then replace 10% of the missing data with constructed data, etc.) until we reach 95% of missing data replaced. For each level adding constructed data, we do 200 replications and average the coefficient, standard error, z-score, and probability for the lower trade dependence variable. Table 6 shows the results of the Monte Carlo simulation for the Gleditsch trade data on the lower trade dependence variable, and Table 7 shows the results for the COW/BKP trade data. The average coefficient values for the Gleditsch data are graphed in Figure 1 and the coefficient values for the COW/BKP data are graphed in Figure 2.

Table 6. Lower Dependence Values from Monte Carlo Runs: Gleditsch Data

Percentage Missing Data Replaced	Average Value for Lower Dependency			
	Coefficient	Std. Err.	Z-score	Prob.
5	-38.125	29.641	-1.282	0.201
10	-37.971	27.960	-1.351	0.177
15	-36.983	26.133	-1.406	0.161
20	-37.246	25.263	-1.464	0.144
25	-36.474	24.101	-1.501	0.135
30	-36.944	23.629	-1.550	0.122
35	-36.176	22.669	-1.580	0.116
40	-36.163	22.175	-1.612	0.109
45	-37.105	22.260	-1.646	0.102
50	-35.946	21.276	-1.662	0.099
55	-36.795	21.561	-1.675	0.098
60	-39.361	22.450	-1.712	0.091
65	-41.082	22.955	-1.752	0.084
70	-42.826	23.834	-1.752	0.084
75	-44.811	24.640	-1.772	0.081
80	-54.039	28.814	-1.822	0.074
85	-56.675	29.920	-1.841	0.072
90	-67.943	34.511	-1.928	0.059
95	-78.144	38.428	-2.019	0.046

Table 7. Lower Dependence Values from Monte Carlo Runs: COW/BKP Data

Percentage Missing Data Replaced	Average Value for Lower Dependency			
	Coefficient	Std. Err.	Z-score	Prob.
5	-38.489	16.538	-2.326	0.020
10	-38.071	16.145	-2.355	0.019
15	-38.517	15.990	-2.405	0.017
20	-37.872	15.572	-2.427	0.016
25	-37.631	15.286	-2.457	0.015
30	-37.217	15.043	-2.468	0.014
35	-37.910	14.995	-2.522	0.012
40	-37.724	14.801	-2.542	0.012
45	-37.972	14.774	-2.562	0.011
50	-37.804	14.715	-2.560	0.012
55	-38.160	14.680	-2.587	0.011
60	-39.373	14.820	-2.646	0.009
65	-38.964	14.748	-2.629	0.010
70	-39.309	14.793	-2.648	0.009
75	-41.951	15.524	-2.689	0.009
80	-43.668	15.900	-2.734	0.007
85	-45.081	16.254	-2.763	0.007
90	-48.337	17.168	-2.803	0.006
95	-54.591	18.850	-2.891	0.004

The overall pattern in both Tables 6 and 7 is that the coefficient values for lower trade dependence gradually get more negative as the amount of estimated trade data increases. Still, note that whereas the probability value does not reach the .05 level until 95% of the missing trade data are replaced with Gleditsch's estimates, lower trade dependence is statistically significant for all levels of estimated COW/BKP data. We see what Allison (2002: 12) says typically occurs with imputed (constructed) data. A slope coefficient will tend to become overestimated. Standard errors are often underestimated using a variety of imputation techniques. Figures 1 (Gleditsch) and 2 (COW/BKP) provide a visual display of how the coefficient values change as the proportion of missing data replaced by estimates is increased. As Figure 1 shows, using the Gleditsch data, there is little change in the average value of the lower dependency coefficient from 5% to about 75%. At that point the absolute value of the coefficient begins to increase rapidly until at 95% of missing values replaced, the coefficient is about twice the size of the coefficient with 5% of missing values replaced. The Monte Carlo study shows that replacing missing trade data with estimates can have a significant effect in both the substantive and statistical sense. However, the Monte Carlo study also suggests that this only happens at the extremes. The same general pattern occurs with the coefficients from the Monte Carlo analysis of the COW/BKP data. But as you can see in Figure 2, the increase in coefficient value as missing data is replaced is smaller.

Figure 1. Gleditsch Average Lower Dependency Coefficient by Percentage of Missing Data Replaced

Figure 2. COW/BKP Average Lower Dependency Coefficient by Percentage of Missing Data Replaced

Conclusions

Our study has demonstrated that when we consider differences in findings about the role that trade dependence plays in the onset of conflict, we need to take the manner in which trade data are generated into account. We feel it is useful to other scholars to see an explicit characterization of the influence of changing the manner in which missing trade data are treated. Taking a step back from our analysis, we see an interesting picture. States with constructed trade figures are less democratic, have lower power capability, and are less developed than those states with observed trade data. These states are thus very different on a set of characteristics we know are important predictors in the international relations literature.¹⁰ Moreover, there were differences in how export and import covaried with the state characteristics we examined. But when we start to include these states with different characteristics (by replacing missing trade data with Gleditsch's constructed trade data), does this have an impact on the relationship between trade dependence and the onset of MIDs? Our conclusion is that it does not change the relationship until a very large proportion of the missing trade data are replaced by estimates. Again, in contrast,

¹⁰ With the exception of region, the state characteristics we use strongly echo the three dichotomies used in Rosenau's classic "Pre-theories" article (Rosenau, 1966). The relative importance of variables in predicting foreign policy behavior is different depending on whether (a) the state has an open or closed political system, (b) the state is large or small, and (c) the state is developed or developing.

when the COW/BKP data are employed, we see that the model estimates may be inflated but not as substantially—and are always statistically significant.

Our analysis of the Gleditsch and COW/BKP data leads us to conclude that replacing missing data with constructed data makes a difference, and the difference appears to be more than simply the impact of increasing the number of cases being analyzed (and its subsequent impact on the statistical significance of coefficients). Different types of states have different chances of having observed, constructed, and missing trade data. This raises the possibility that one's decisions on missing trade data will not only affect the relationship between trade and conflict, but may affect other relationships as well. Clearly, the problem of list-wise deletion can be quite severe, especially in regard to economic data where some data sources can be unreliable and when the same states that are prone to missing trade data are similarly missing on other covariates.

This leads to a predictable question. Which trade data set is better? The answer to that question could not and should not be stated in an absolute manner because it depends on one's research question. Our goal is not to decide which one of these data sets is "better" than the other. We think that is neither possible nor appropriate; however, we believe our efforts will help researchers to use the trade data that are more appropriate for the questions they wish to ask. We also do not believe that differences in findings in this field of research can be attributed solely to the manner in which different researchers have treated missing trade data. Recent studies clearly show that other decisions—decisions about indicators, time frames, and techniques of analysis—matter. We do believe, however, that decisions about missing trade data also have consequences, and we urge our colleagues to take these differences into consideration when contemplating which trade data to employ. Each of these data sets is still likely to contain observations that could be corrected, as is the case with most data sets. With that said, we implore, as others have done in the past, scholars to know their data.

It is reasonable and beneficial that some trade data are built so as to reduce some potential sample bias and increase estimation efficiency in our studies of interstate conflict. The problem of missing trade data will vary with differences in samples and for what purposes they are used. If one were to use trade data as control variables, then perhaps this issue would be slightly less important. If one is going to study trade in particular, then it would pay to make informed choices based on what types of estimated data one finds acceptable. For example, when trade data are missing for one state in a dyad, filling in the missing data could produce systematic error such as providing artificially higher levels of symmetry in trade asymmetry variables. Perhaps this is a bit of a test for risk acceptant versus risk adverse behavior, but it does come down to calculating how much potential error one is willing to inject into one's study through the adoption of assumptions underlying imputation methods versus potentially constrained variance for trade dependence and sample bias. However, the magnitude of this problem will vary from study to study because the negative implications of missing trade data are like multicollinearity in that it is a sample specific loss of information problem for statistical estimation.

References

- Allison, Paul D. 2002. *Missing Data*. Thousand Oaks, CA: Sage University Papers Series on Quantitative Applications in the Social Sciences.
- Barbieri, Katherine. 1995. Economic interdependence and militarized conflict, 1870–1985. PhD dissertation. State University of New York at Binghamton.
- Barbieri, Katherine. 1996. Economic interdependence: A path to peace or a source of interstate conflict? *Journal of Peace Research* 33(1): 29–49.
- Barbieri, Katherine. 1998. *International Trade Database (1870–1992)*. Version 1.1 (http://pss.la.psu.edu/TRD_DATA.htm).
- Barbieri, Katherine. 2002. *The Liberal Illusion: Does Trade Promote Peace?* Ann Arbor, MI: University of Michigan Press.
- Barbieri, Katherine, and Richard Allan Peters II. 2003. Measure for mis-measure: A response to Gartzke and Li. *Journal of Peace Research* 40(6): 713–719.
- Barbieri, Katherine, Omar Keshk, and Brian Pollins. 2008. Correlates of War Project Trade Data Set Codebook, Version 2.01 (<http://correlatesofwar.org>).
- Barbieri, Katherine, Omar Keshk, and Brian Pollins. 2009. Trading data: Evaluating our assumptions and coding rules. *Conflict Management and Peace Science* 26(5): 471–491.
- Beck, Nathaniel, and Jonathan N. Katz. 2001. Throwing out the baby with the bath water: A comment on Green, Kim, and Yoon. *International Organization* 55(2): 487–495.
- Beck, Nathaniel, Jonathan N. 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): 1260–1288.
- Bennett, D. Scott, and Allan Stam. 2000. *EUGene: A Conceptual Manual*. *International Interactions* 26: 179–204.
- Dafoe, Allan. 2008. Democracy still matters: The risks of sample censoring, and cross-sectional and temporal controls. Manuscript.
- Gartzke, Erik. 2007. The capitalist peace. *American Journal of Political Science* 51(1): 166–191.
- Gartzke, Erik, and Quan Li. 2003a. Measure for measure: Concept operationalization and the trade independence–conflict debate. *Journal of Peace Research* 40(5): 553–571.
- Gartzke, Erik, and Quan Li. 2003b. All's well that ends well: A reply to Oneal, Barbieri, and Peters. *Journal of Peace Research* 40(6): 727–732.
- Gartzke, Erik, and Quan Li. 2005. Mistaken identity: A reply to Hegre. *Conflict Management and Peace Science* 22(3): 225–233.
- Gelpi, Christopher F., and Joseph M. Grieco. 2008. Democracy, interdependence, and the sources of the liberal peace. *Journal of Peace Research* 45(1): 17–36.
- Gleditsch, Kristian Skrede. 2002. Expanded trade and GDP data. *Journal of Conflict Resolution* 46(5): 712–724.
- Gleditsch, Kristian Skrede. 2004. Expanded trade and GDP data (<http://weber.ucsd.edu/~kgledits/exptradegdp.html>, accessed 18 October 2005).
- Gleditsch, Kristian Skrede, and Michael Ward. 1999. A revised list of independent states since 1816. *International Interactions* 25(4): 393–413.
- Green, Donald P., Soo Yeon Kim, and David H. Yoon. 2001. Dirty pool. *International Organization* 55(2): 441–468.
- Gujarati, Damodar N. 2003. *Basic Econometrics*, 4th edn. New York: McGraw Hill.
- Hegre, Håvard. 2005a. Identifying how trade matters in empirical studies of interstate conflict. *Conflict Management and Peace Science* 22(3): 217–224.
- Hegre, Håvard. 2005b. Identifying how trade matters: A response to Gartzke and Li. *Conflict Management and Peace Science* 22(3): 235–237.

- Keshk, Omar, Brian Pollins, and Rafael Reuveny. 2004. Trade still follows the flag: The primacy of politics in a simultaneous model of interdependence and armed conflict. *Journal of Politics* 66(3): 1155–1179.
- Keshk, Omar, Rafael Reuveny, and Brian Pollins. 2010. Trade and conflict: Proximity, country size, and measures. *Conflict Management and Peace Science* 27(1): 3–27.
- King, Gary. 2001. Proper nouns and methodological propriety: Pooling dyads in international relations data. *International Organization* 55(2): 497–507.
- King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review* 95(1): 49–69.
- Lewis-Beck, Michael S. 1980. *Applied Regression: An Introduction*. Newbury Park, CA: Sage University Papers Series on Quantitative Applications in the Social Sciences.
- Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff. 2000. Free to trade: Democracies, autocracies, and international trade. *American Political Science Review* 94(2): 305–321.
- Oneal, John R. 2003. Measuring interdependence and its pacific benefits: A reply to Gartzke and Li. *Journal of Peace Research* 40(6): 721–725.
- Oneal, John R., and Bruce M. Russett. 1997. The classical liberals were right: Democracy, interdependence, and conflict, 1950–85. *International Studies Quarterly* 41(2): 267–295.
- Oneal, John R., and Bruce M. Russett. 1999a. Assessing the liberal peace with alternative specifications: Trade still reduces conflict. *Journal of Peace Research* 36(4): 423–442.
- Oneal, John R., and Bruce M. Russett. 1999b. The Kantian peace: The pacific benefits of democracy, interdependence, and international organizations, 1885–1992. *World Politics* 52(1): 1–37.
- Oneal, John R., and Bruce M. Russett. 2001. Clear and clean: The fixed effects of the liberal peace. *International Organization* 55(2): 469–485.
- Oneal, John R., Frances H. Oneal, Zeev Maoz, and Bruce M. Russett. 1996. The liberal peace: Interdependence, democracy, and international conflict, 1950–1986. *Journal of Peace Research* 33(1): 11–28.
- Oneal, John R., Bruce M. Russett, and Michael L. Berbaum. 2003. Causes of peace: Democracy, interdependence, and international organizations, 1885–1992. *International Studies Quarterly* 47(3): 371–393.
- Rosenau, James. 1966. Pre-theories and theories of foreign policy. In *Approaches to Comparative and International Politics*, ed. R. Barry Farrell, pp. 27–92. Evanston, IL: Northwestern University Press.
- Russett, Bruce M., and John R. Oneal. 2001. *Triangulating Peace: Democracy, Interdependence, and International Organizations*. New York: W.W. Norton.

CHARLES R. BOEHMER is an Associate Professor and Director of Graduate Studies in the Department of Political Science at the University of Texas at El Paso. His current research interests include international conflict, civil conflict, economic influences on politics, and international organization.

BERNADETTE M.E. JUNGBLUT is the Director of Assessment and Retention at West Virginia University. Her current research interests include examining risk factors for student attrition and evaluating the impact of different intervention strategies.

RICHARD J. STOLL is the Albert Thomas Professor of Political Science at Rice University. He is also a Baker Institute Scholar at the James A. Baker III Institute for Public Policy at Rice University. His research focuses on the causes of international conflict, national security policy, and public opinion and foreign policy.

Appendix

Case Structures for Trade Figures and State Characteristics Analysis

Gleditsch Trade Data

The first case structure to be analyzed is derived from Gleditsch's "expdata.asc" trade data set. As noted above, each case in that data set is a non-directed dyad with multiple trade data entries. The relevant data for each record is as follows (note: this is not the exact format, but is rearranged and other variables omitted to simplify discussion):

Year StateA Aexport_toB Aimport_fromB StateB Bexport_toA Bimport_fromA

The second case structure to be analyzed is the COW/BKP data, which includes variables measuring the flow of imports between states A and B:

Year StateA Flow1(imports of A from B) StateB Flow2(imports of B from A)

There is an additional record for state A with every other state in the Gleditsch-Ward system in that year. We need to organize these data so that we can conduct our analysis.

First, for the analysis of state characteristics, we do not use trade figures. We instead use the associated variable for each trade figure that indicates whether the trade figure is observed, constructed, or missing. We then take each dyadic record (such as the one in the example above) and convert it into two state-level records:

Year StateA Aexport_toB Aimport_fromB Otherstate (this is StateB)

Year StateB Bexport_toA Bimport_fromA Otherstate (this is StateA)

We then add the set of dummy variables for the type of trade data to each record (note since there are both import and export trade data on a single record, there is a set of three dummies for each trade figure).

The COW/BKP Data

Analogously, the set-up for the BKP data is:

Year StateA Flow1 Otherstate (this is StateB)

Year StateB Flow2 Otherstate (this is StateA)

We then add the state characteristics data to each record. For the first record in the above example, the state characteristics are for State A, and for the second record, the state characteristics are for State B. For analyses of state characteristics that are continuous, such as democracy or CINC, we create three dummy variables, one for each category of trade figure. We then regress the state characteristic on the three dummy variables (omitting the constant). For the rest of the state characteristics, we report contingency tables, comparing the three-way characterization of the type

of trade data with the state characteristic. Note that we conduct different analyses for each state's trade export data and its import data and, analogously, for its Flow 1 and Flow 2 data.

Table A1. Contingency Table Results for Analysis of Major Powers

<i>Major Power vs. Export Data Type</i>				
<i>Export Data Center</i>				
<i>Major Power</i>	<i>Missing</i>	<i>Constructed</i>	<i>Observed</i>	<i>Total</i>
No	222,960 99.25%	303,361 98.52%	535,922 94.13%	1,062,243
Yes	1,675 0.75%	4,548 1.48%	33,393 5.87%	39,616
Total	224,635	307,909	569,315	1,101,859
<i>Major Power vs. Source1 Data Type</i>				
<i>Source1 Data Type</i>				
<i>Major Power</i>	<i>Missing</i>	<i>Constructed</i>	<i>Observed</i>	<i>Total</i>
No	291,655 98.47%	92,853 95.15%	562,727 93.51%	947,235
Yes	4,545 1.53%	4,738 4.85%	39,054 6.49%	48,337
Total	296,200	97,591	601,781	995,572
<i>Major Power vs. Import Data Type</i>				
<i>Import Data Type</i>				
<i>Major Power</i>	<i>Missing</i>	<i>Constructed</i>	<i>Observed</i>	<i>Total</i>
No	222,960 99.25%	375,357 97.12%	463,926 94.53%	1,062,243
Yes	1,675 0.75%	11,113 2.88%	26,828 5.43%	39,616
Total	224,635	386,410	490,754	1,101,859
<i>Major Power vs. Source2 Data Type</i>				
<i>Source2 Data Type</i>				
<i>Major Power</i>	<i>Missing</i>	<i>Constructed</i>	<i>Observed</i>	<i>Total</i>
No	296,837 98.78%	116,847 95.62%	533,551 93.14%	947,235
Yes	3,681 1.22%	5,353 4.38%	39,303 6.86%	48,337
Total	300,518	122,200	572,854	995,572

Table A2. Trade Data Type by Region: Gleditsch Data

Region	Export Data Type			Total
	Missing	Constructed	Observed	
Europe	29,011 11.91%	44,345 18.21%	170,222 69.88%	243,578
Middle East	20,869 15.06%	45,321 32.71%	72,354 52.22%	138,544
Africa	79,543 28.77%	98,360 35.58%	98,750 35.65%	276,473
Asia	50,463 23.34%	53,419 24.70%	112,366 51.96%	216,248
Americas	44,749 19.71%	66,464 29.28%	115,803 51.01%	227,016
Total	224,635 20.39%	307,909 27.94%	569,315 51.67%	1,101,859

Region	Import Data Type			Total
	Missing	Constructed	Observed	
Europe	29,011 11.91%	75,558 31.02%	139,009 57.07%	243,578
Middle East	20,869 15.06%	50,289 36.30%	67,386 48.64%	138,544
Africa	79,543 28.77%	100,375 36.31%	96,555 34.92%	276,473
Asia	50,463 23.34%	83,709 38.71%	82,076 37.95%	216,248
Americas	44,749 19.71%	76,539 33.72%	105,728 46.57%	227,016
Total	224,635 20.39%	386,470 35.07%	490,754 44.54%	1,101,859

Table A3. Trade Data Type by Region: COW/BKP Data

Region	Export Data Type			Total
	Missing	Constructed	Observed	
Europe	62,190 20.89%	29,204 9.81%	206,272 69.30%	297,666
Middle East	18,248 27.51%	7,283 10.98%	40,795 61.51%	66,326
Africa	87,256 38.59%	26,082 11.53%	112,779 49.88%	226,117
Asia	13,873 36.23%	4,998 13.05%	19,425 50.72%	38,296
Americas	114,633 31.22%	30,024 8.18%	222,510 60.60%	367,167
Total	296,200 29.75%	97,591 9.80%	601,781 60.45%	995,572

Table A3. (Continued)

Region	Import Data Type			Total
	Missing	Constructed	Observed	
Europe	53,823 18.08%	41,871 14.07%	201,972 67.85%	297,666
Middle East	20,295 30.60%	8,185 12.34%	37,846 57.06%	66,326
Africa	97,421 43.08%	25,263 11.17%	103,433 45.74%	226,117
Asia	13,724 35.84%	3,362 8.78%	21,210 55.38%	38,296
Americas	115,255 31.39%	43,519 11.85%	208,393 56.76%	367,167
Total	300,518 30.19%	122,200 12.27%	572,854 57.54%	995,572

Table A4. Gleditsch's Analysis Predicting Onset of Militarized Interstate Disputes from Economic Dependence and Distance

	Maximum Missing Data	Maximum Estimated Data
Constant	2.113 (0.217)	3.681 (0.219)
Trade Dependence (Weak Link)	-33.600 (9.261)	-49.317 (0.787)
Natural Log of Distance Between Capital Cities	-0.893 (0.029)	-1.131 (0.018)
N	92,096	394,057
LR χ^2 (df = 2)	914.19	3799.63