Conflict Management and Peace Science

[DOI:10.1177/0738894209343887]
Vol 26(5): 471–491

cmps

Trading Data:

Evaluating our Assumptions and Coding Rules[1]

KATHERINE BARBIERI
University of South Carolina, USA

OMAR M. G. KESHK
The Ohio State University, USA

BRIAN M. POLLINS
The Ohio State University, USA

Some scholars have rushed to judgment about the nature of the relationship between trade and conflict, making strong assumptions about the data upon which their conclusions rest. In this paper, we test these assumptions, showing that they are often not warranted and, thus, pose threats to many of our conclusions about trade's impact on conflict. We discuss official trade statistics; the treatment of missing trade data; and problems with some decision rules being adopted within our research community. We introduce the new Correlates of War (COW) Trade Data Set; discuss the rationale behind our coding decisions; and compare this data set with other sets. The end result is a series of findings that should help our field advance its understanding of the often difficult issue of trade's relationship with international conflict.

KEYWORDS: COW Trade Data Set; IMF data; imputed data; militarized interstate disputes; missing data; trade; trade and conflict; trade data; trade statistics

Introduction

The emergence of the sub-field of international political economy (IPE) in the 1970s reframed the attention of a discipline that had largely ignored economic variables. Suddenly, economics, including relations of dependence and interdependence, became central to our understanding of peace, security, and justice. Despite the theoretical interest, scholars seeking to test hypotheses about economics and conflict faced

[1] We would like to thank John Geer, Ranan Kuperman, Glenn Palmer, Paul Paust , and James Lee Ray for their helpful comments.

numerous hurdles in acquiring quality data. The field, which had previously paid too much attention to Morgenthau's *Politics Among Nations* (1948), now paid too little heed to Morgenstern's (1963) *On the Accuracy of Economic Observations*.

Too little consideration was given to the accuracy of economic data, the expertise of the analysts generating it, coding decisions, or the rationale for combining different data types into aggregate variables.[2] Unfortunately, these problems continue to plague our research community, particularly in the area of international trade statistics.[3] In this paper, we discuss some of the general problems associated with official trade statistics;[4] the treatment of missing trade data; and problems with several decision rules being employed in the field. Next, we introduce the new Correlates of War (COW) Trade Data Set; discuss the rationale behind our coding decisions; and compare these data with other available evidence. Our findings call into question some of the conclusions that previous work has made about trade and conflict.

Common Problems with Official Trade Statistics

Scholars interested in accurately depicting trading relationships confront a series of challenges. We identify and discuss what we believe are three of the most serious problems scholars face when using official trade statistics: (1) one or both states in a dyad file inaccurate reports; (2) two states report dyadic trade, but the values reported differ considerably by state; or (3) one or both countries fail to report trade. These are not mutually exclusive categories. For example, if the International Monetary Fund (IMF) trade data set suggests that one state does not trade with another state (i.e. a zero value is attached to the dyad), the report may be (1) accurate or inaccurate; (2) similarly reported by both states in the dyad; or (3) the product of IMF adjustments to missing national reports. Scholars who observe the zero value are unlikely to know or question the origin and accuracy of the IMF's figure. We will discuss, in turn, the causes and consequences of several critical trade data problems. Later, we discuss how we deal with these problems and how others handle them.

Inaccurate Trade Reports

Inaccurate trade reports may result from deliberate or non-deliberate acts by governments who compile information; publish trade records; and submit these to international organizations. The international organization itself might also introduce errors. It is often difficult to compare national trade reports and assess their

---

[2] For example, many scholars created ratio dependence measures where they relied on figures that had the dollar value of the denominator (GDP) adjusted with a PPP index, but no index was applied to the numerator (Trade).

[3] Fully reviewing the trade–conflict literature is beyond the scope of this study. See Schneider et al., 2003; and Mansfield and Pollins, 2003 for thorough reviews.

[4] For problems with trade statistics, see Bhagwati, 1964, 1967; Bhagwati and Srinivasan, 1981; Ely, 1961; Morgenstern, 1963; Pak and Zdanowicz, 1994a,b; Pak et al., 2003; Sheikh, 1974; Yeats, 1978, 1990; de Wulf, 1981.

accuracy, since reporting procedures vary across countries. Despite the IMF's efforts to standardize reporting practices, differences remain among states and across time. Researchers should keep this in mind when employing historical time series data. For example, in the past, most states reported *special trade* (direct trade between the point of origin and point of destination), but a few states reported only *general trade* (special trade plus transit trade—goods that flow through an area on their way to their final destination). Scholars typically ignore the distinction between special and general trade. Yet, these figures may be significantly different and they may also provide insight into the role of transit traders in global economics and politics.

For example, transit traders may serve as a conduit through which hostile states trade and may provide a means for circumventing sanctions, quotas, and other trade laws. This and other types of illegal trade may result in inaccurate trade figures. Advances by scholars who research and estimate illegal trade are seldom incorporated into traditional trade research. Although we do not solve this problem here, we believe it is important to consider what is missing in official trade figures and to be cautious when making assumptions about trade data. Even when trade is legal, it may be difficult to determine whether national trade statistics are accurate, whether reporting errors are deliberate, and whether firms, states, or other actors are responsible for the inaccuracies. Scholars typically assume that developed states have more advanced techniques than developing states to compile data, monitor firm activities, and ascertain the origin and destination of traded goods. Certainly, one source of inaccurate reporting is suboptimal recording and monitoring systems. There are great variations across countries in terms of customs administrations, law enforcement, and border patrol. Here, the government might wish to report trade accurately, but be unable to do so.

False reports may also be the product of politically motivated efforts to distort trade figures. A state might inflate its trade figures with a given partner when confronted with pressure to open its markets to goods from that country. States might also inflate trade figures when they wish to project an image of economic growth (e.g. for domestic audiences or to attract foreign investment). Transit traders might also distort trade figures to hide sanctions-busting activities. Even when trade is legal, both states and firms may wish to avoid public image problems that could come with trading with pariah states. In addition, governments might not report all trade in military or sensitive goods (IMF, 2008). Clearly, there are various reasons why states might hide legal and illegal trade.

In some cases, governments may not be aware that trade exists between their citizens and those other countries, or they may recognize the existence of trade ties but be unable to record or prevent them. For example, the United States is aware that illegal products cross its borders daily and may even know some actors involved in this trade, but it is unable to stop illegal trade or record its value accurately. Even when states develop estimates of illegal trade and transnational ties, based upon seizures and information from violators caught in illegal activities, these estimates are typically excluded from official trade statistics.[5]

[5] According to the UN, some nations include illegal trade in their national trade statistics, but the UN gives no explanation for how the estimates are derived (UN, 2008).

Firms may also have economic and political incentives to mischaracterize real trade values, which could result in inaccurate national reports. When economic and political incentives increase the desirability for firms to mischaracterize point of origin, destination, trade values, etc., national and dyadic trade figures may be inaccurate. As mentioned, firms may be concerned with being publicly linked to repressive regimes. They may also seek to avoid high tax rates, quotas, and other trade restrictions. If firms consistently engage in transfer pricing, where trade is underreported with states that have high tax rates and underreported with those that have low rates, there will be a systematic bias in the patterns of trade reported.

Understanding when and why states or firms might manipulate trade figures is interesting for a number of reasons—the most important being that inaccurate figures undermine our ability to understand trading relationships and their causes and consequences in international politics. Clearly, the problems of inaccurate reporting are difficult to overcome, but understanding when and where problems may exist is a necessary first step to solving them.

One way to assess the accuracy of a given nation's trade figures is to compare these with its trading partners' figures. This may reveal whether one state tends to over or underreport trade, relative to its partners, or whether it fails to report trade with a given state or category of states (e.g. Israeli trade with its neighbors). This certainly does not solve all of our problems, but could highlight the challenges facing researchers using dyadic trade figures. This brings us to the second category of trade problems—inconsistent trade reports between partners.

Discrepant Partner Reports

Many of the strategies that scholars develop to address problems of underreporting or nonreporting trade are based upon the assumption that two countries tend to provide similar depictions of the same trading relationship. However, trade experts have long recognized the problem of discrepant partner reports (see Morgenstern, 1963). In addition, we should not expect an importer and an exporter to report identical trade values, since the importing state typically reports the value of goods traded in c.i.f. (cost, insurance, and freight) values, which includes transportation costs, while the exporter reports them in f.o.b. (free on board) values, which excludes transport costs.

The IMF outlines several reasons for differences in partner reports. These include: differences in classification schemes, recording times, valuations, coverage, and processing errors as causes of inconsistent reports (IMF, 2007). Differences are also tied to "inconsistent currency version, evasion, anti-evasion procedures, values not known at the time of consignment, and differences in treatment of particular costs or procedures for assessment" (IMF, 2008). Among the things the IMF (2008) describes as coverage problems are "shipments to and from free-trade zones and bonded warehouses, exclusion of military and other confidential items and government goods, value thresholds for customs registration of shipments, returned goods, and other goods missed by customs (or surveys)."

The IMF estimates that the difference in values between the importer and exporter is approximately ten percent, as a result of importers reporting trade in c.i.f. values,

while exporters typically use f.o.b. values.[6] F.o.b. values include the price of goods at the port of origin and include domestic transportation fees. The c.i.f. value, on the other hand, is assessed at the destination port and includes both the f.o.b. price components and the costs of international transportation and insurance. At times, the IMF adjusts trade values up or down by ten percent in order to transform one state's reported c.i.f. values into f.o.b. values or vice versa or when the IMF decides to use one state's reported trade values to substitute for its partner's missing trade report.

The IMF's adjustment practices and their ten percent rule pose problems for scholars interested in the degree to which trade partner reports coincide. First, scholars are typically unaware if and when the IMF adjusts data before publishing it. They may assume two states provide consistent dyadic trade reports, when the reports are simply a product of IMF adjustments. Certainly, one would expect two states' trade reports to be highly correlated if they are based upon the same data that come from one state.

The IMF's ten percent rule may be particularly problematic for conflict researchers, since it could grossly undervalue transportation fees and mask trade ties that exist during conflicts. Using an exporter's trade report to replace a missing importer's report assumes that the price of goods remains relatively constant from the time goods leave the port of origin until they reach their destination, but delivery delays and other factors could increase the gap between the exporter's and importer's price. Periods of intense conflict and instability within a state, region, or relationship are likely to result in delivery delays, shortages, increases in insurance premiums, and higher transportation costs. Data are also less likely to be systematically collected or reported during periods of intense internal or external conflict. There may be political reasons why states or firms want to hide or distort trade ties or delay the release of trade reports.

In addition, new security measures instituted after 11 September 2001 have placed added burdens on shippers.[7] The IMF's ten percent figure may be less accurate today than in the past and less accurate in conflict zones than peaceful areas. The IMF's decision rule cleanses trade data of important information for conflict researchers and may mask the realities we hope to investigate. For example, variations in insurance and shipping costs across space and time could provide valuable information about changing political and economic conditions, rising tensions, and perceived risk in international business.

The question is how often does one state report dyadic trade when its partner does not? The problem here is that the IMF does not provide adequate information to allow us to know when and how often the IMF uses the ten percent rule. We only know when it does not happen—when we have a dyad where only one state reports trade in the relationship. We found that in at least 14% of our unadjusted IMF observations, only one of two states in the relationship reports trade. In the

---

[6] The following countries report their imports in f.o.b.: Australia, Bermuda, Brazil, Canada, Czechoslovakia, Dominican Republic, Mexico, Papua New Guinea, Paraguay, Peru, Romania, South Africa, and Zimbabwe. The IMF adjusts these figures and reports them in c.i.f.

[7] See Barbieri and Pathak (2007) for a discussion of post-9/11 trade security regulations.

majority of cases—64% of observations—either both states report trade or one state reports trade and the IMF uses that figure to replace the other state's missing report. In approximately 22% of our cases, both states in a dyad do not report trade in their relationship.

Another important question for researchers is how great the variation is in most partner reports and what impact this might have on our research. Once again, we confront the problem of not knowing the impact of IMF adjustments on our figures. Yet, we measured the magnitude of inconsistencies in partner reports, by dividing the larger value of reported trade in a given dyadic flow by the smaller reported trade value. We found that in approximately 73% of cases, the importer and exporter reports differed by more than ten percent. The differences were ten percent or less in approximately 26% of partner reports. As mentioned above, some of the similar values are a function of IMF data adjustments, so we are likely to underestimate the differences in reports.

Scholars must consider how they might want to portray a trading relationship when dramatic differences exist in partner reports. There may be times when we want to understand why differences exist between state reports. In reality, the dyad has one trading relationship that we want to depict. Our desire to create one operational measure of dyadic trade affects the decision rules we adopt to produce our data set.

Missing Trade Reports

One of the most difficult challenges occurs when one state fails to report its trade with another. This produces what we typically describe as "missing data." As mentioned, approximately 14% of all IMF observations are situations in which there are no trade values reported for one side of a dyadic relationship, and there are no reports for either state in the dyad in 22% of IMF cases. Researchers must deal with the question of whether we should assume missing data indicate an absence of trade or whether the missing values conform to a particular trend in the data series. Given that we know there are instances in which one state reports trade and the other does not, we know missing data does not always mean trade is absent.

If we know trade does not exist, we could replace missing data with zero values. However, it is difficult to conclude that no trade exists between two states, unless we dig deeper into why the data might be missing. In the past, the IMF frequently reported zero trade between states when the value of dyadic trade fell below a certain minimum threshold. The reported value was available in print versions of the IMF's trade statistics and from other sources. Yet, most scholars opted not to look beyond the zero values; some went even further and assumed missing data indicated zero trade. In these and other cases, we opted to collect data, rather than generate it artificially. Fortunately, the IMF has altered its practice of replacing minimum values with zero values in its electronic data releases. Yet, many scholars continue to make assumptions about missing data that could threaten the reliability of our data sets.

For example, one must consider how to deal with gaps in a data series. Suppose two states appear to have an upward trend in their trade over time but the data series have several years of missing data. Should we simply assume the missing years conform to the same trend and interpolate to create data to replace missing data?

Suppose these two states stopped trading during those years or stopped reporting it, because something usual happened. The states may have even fought a war; one may have invaded and occupied the other; and they may have resumed trade and even increased it afterwards. Data may be missing for many reasons—some of them tied to conflict. If we simply manufacture data based upon flawed assumptions, we threaten our ability to make accurate inferences about trade and conflict relationships.

Clearly, conflict scholars must make a number of decisions about how to handle missing and other data problems. In creating the Correlates of War Trade Data Set, we considered the issues described above and developed coding decisions designed to minimize measurement errors. We recognize that any large data set is prone to errors, but we believe we have gone further than most in reducing errors that might be particularly problematic for conflict scholars.

The Correlates of War Bilateral Trade Data

The COW Trade Data Set, Version 2.0, builds upon several of the authors' earlier projects, including BKP Version 1.0 (see Barbieri et al., 2003) and Barbieri's International Trade Data Base, Version 1.0 (Barbieri, 2002, Appendix A).[8] The COW Trade Data Set includes dyadic and national trade figures for state system members (COW Project, 2008) for the period 1870–2006. The set includes three files: (1) dyadic trade statistics; (2) national trade statistics; and (3) a codebook that describes the variables and details about the collection procedures. The majority of the post-WWII data for the COW Trade Data Set were obtained from the IMF's *Direction of Trade Statistics* (DOTS) (IMF CD-ROM, 2007), while the pre-WWII data are from Barbieri (2002).

Today, the IMF is the most commonly used source of trade data for scholars, so we focus our attention primarily on these data. The IMF obtains data from its members. As its membership has expanded over time, data from previously excluded countries and regions (e.g. Eastern Europe in the Cold War) have been updated. This provides us with a better picture of trading relationships than we previously had. The quality and precision of the IMF data have also improved over time. The IMF's monthly data releases include corrections to previous releases (e.g. typographical, decimal positioning, dropped numbers, member updates). The data errors, at times, result in trade figures that differ from their correct value by a factor of one thousand or more.

Over time, the IMF has employed a variety of decision rules to handle data problems. Unfortunately, the IMF does not make it clear when it applies its rules. For example, it replaces some missing data with partner reports, but not others. As we discuss below, some of the coding rules may be problematic for conflict scholars. Moreover, moving from IMF electronic data releases to a user-friendly dyadic format requires additional coding decisions and considerable work.

To construct the COW Trade Data Set, we first constructed a data matrix of IMF partner pairs, for the period 1948–2006, using the IMF's *DOTS* CD-ROM (2006, 2007). The IMF partner list includes state and non-state actors and aggregations of states by

[8] BKP Version 1.0 used the IMF's *DOTS* from ICPSR Study No. 7628, Computer File 1996. See the IMF (1993, 2007, 2008) for information about their trade data.

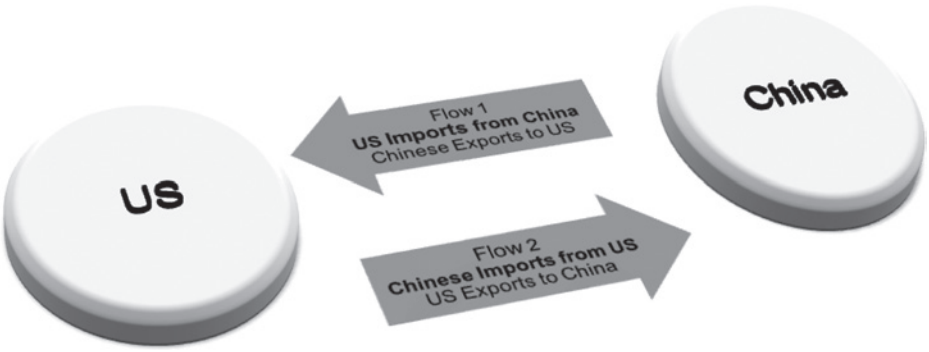

Figure 1. Simplified Version of the Dyadic Trading Relationship between the US and China

region, association, level of development, and other groups. We narrowed our focus to dyads composed of state partners and made changes to the IMF state list and dates, so it would match the COW state system list and dates. After producing the IMF trade data matrix, we applied a series of coding rules to move to the next phase of our data construction process. To understand our strategy for constructing the dyadic trade data, it helps to consider an example. Figure 1 depicts a simplified version of the trading relationship between the US and China.

In the US–China dyad and any other bilateral trading relationship, goods flow in two directions (e.g. East–West, West–East, etc.). The exporter and importer report the value of the goods traded. Thus, we have two elements to consider in our data. One is the direction of the trade flow and the other is the source of the report (importer versus exporter). Imagine a vessel docked in a Chinese port, loaded with containers of Chinese goods destined for the US market. When the ship leaves its dock, China records the value of these goods as Chinese exports to the US; when they arrive at a US port, US Customs Officials report the value of these goods as US imports from China. At the end of each year, the US and China each provide the IMF with the aggregate value of all goods flowing between their two states.

In Figure 1, we see two directional trade flows, Flow 1 and Flow 2. For each, we have an importer and an exporter report. This gives us four trade values: (1) Flow 1, importer report; (2) Flow 1, exporter report; (3) Flow 2, importer report; and (4) Flow 2, exporter report. Again, this is clearly a simplification of the process, but it should illuminate the issues we confront.

Scholars might reasonably conclude that the importer and exporter trade reports would be roughly equal and that we could rely upon either state for our trade data. Yet, partner reports are not always equal, as we discussed above. Moreover, we should not expect them to be identical. As mentioned above, most states report imports in c.i.f. values (i.e. cost, insurance, freight) and exports in f.o.b. values (i.e. free on board); c.i.f. values should be larger than f.o.b. values. For example, if China reports exports of one million dollars' worth of goods to the US, using f.o.b. figures, the US, as importer, should add shipping and insurance costs to the million dollar figure to produce the c.i.f. import value. The US's import value for Flow 1 should be higher than China's export report, all else being equal.

Ideally, a researcher would have data from the importer and exporter for each directional flow (Flow 1 and Flow 2), but this ideal is not always realized. Even if four data points are available for a given relationships, it may not be optimal to use all of them. When possible, we relied upon the importer's trade values, rather than the exporter's figures, because states have a greater tendency to monitor goods coming into their country, rather than leaving it. For example, states tend to tax imports more often than exports, and they tend to be more concerned that guns will flow into, rather than out of, their country. By using the importer's report to characterize each directional flow, our dyad consists of information from both states. This should help us reduce, although not eliminate, measurement error that could arise if one state has a systematic bias in reporting.

In constructing the COW Trade Data Set, we started with a list of directed dyads and used data from the importing state (the US, above) to characterize Flow 1 and the importing state (China) to characterize Flow 2. We used the IMF data to fill in all import cells for State 1, since State 2 is always the exporter in the directed dyad structure.[9] When one of the importer's reports was missing, we used the exporter's trade report to replace missing and zero values. Within the directed dyadic format, we really have only Flow 1 listed, along with the importer and exporter values, since the US–China and China–US are considered two different dyads. These rows were merged when we moved to the non-directed dyad format, so we had only the China–US dyad and Flow 1 and Flow 2, where the US and China alternated positions as importer and exporter.

Once we had filled in all possible data cells using current IMF data, we turned to alternative data sources to replace missing values. We started with the BKP data set, which includes information from earlier IMF data tapes (1996). Next, we used Barbieri's trade data; this set includes IMF, non-IMF, and pre-WWII data. In several special cases (Taiwan, Belgium–Luxembourg, and China), we used additional sources or made adjustments to aggregate or disaggregated trade figures.

In the case of Taiwan, the IMF does not provide trade data, so we compiled data from the Republic of China's government websites. In the case of Belgium and Luxembourg, the IMF reported one aggregate value for these states until 1996, after which time data are separated by state. We calculated disaggregated trade data for the pre-1996 period, based on the relative size of Luxembourg and Belgium's GDPs. To do this, we obtained annual GDP data for Belgium and Luxembourg (World Bank, 2005) and generated an annual ratio value of the smaller to higher GDP values (i.e. Luxembourg to Belgium). We multiplied Luxembourg's dyadic trade figures by this ratio value and multiplied Belgium's trade figures by one minus the ratio value. While not an ideal solution, we recommend using the disaggregated values; but we also provide the IMF's original figures for interested scholars. [10]

---

[9]In the COW Trade Data Set, states are arranged according to their COW country codes (CCode), where CCode1 is the state with the lower value and CCode 2 is the higher value.

[10]Another strategy would be using post-1996 dyadic trade figures for these two states to generate dyad specific ratios that are applied to the pre-1996 figures.

For China, we confront the opposite problem—disaggregated, rather than aggregated trade reports. Here, the IMF continues to report separate trade values for China, Macao, and Hong Kong despite the fact that these areas were unified after 1998. Since this might seriously undervalue China's trade, we added Macao and Hong Kong to China's reported trade values after 1998. The average difference between China's dyadic trade values before and after Macao and Hong Kong are added is approximately 27% for the period 1999–2006. Once again, we include the IMF's original China figures in a separate column.

Using all of the decision rules described above, the COW Trade Data Set has approximately 19% more trade values than the original IMF data base for the post-WWII period. We also added data for the pre-WWII period (1870–1947). We summarize our data sources and procedures in Table 1. We still have a great deal of missing data, which we code as missing. Our decision to code missing values as such differs from scholars who assume missing data signifies an absence of trade between states or those who assume that dyadic trade continues according to a linear trend, even when values are missing, and that we could apply this trend to generate values for missing data. We believe these assumptions and several others being adopted in our research community are often faulty and could threaten our quest for greater scientific understanding of the causes and consequences of trade. Next, we turn to a discussion of some of these decision rules for handling missing data.

Decision Rules for Missing Data

Scholars have devised strategies to handle the missing data problem. We believe some of these are more problematic than others for conflict researchers. We focus, in particular, on the decision rules adopted by Gleditsch (2002), since his data are frequently employed in the field. Oneal and Russett's (1997) data sets are also widely used and suffer from many of the same problems. The most recent versions of the data sets discussed here contain far fewer problems than earlier versions; unfortunately, most published studies are based on the earlier data sets.[11]

Gleditsch's (2002) database provides trade figures for imports and exports and relies primarily on IMF data. Gleditsch first replaces missing or zero trade values with IMF export data and then uses the World Export Data set (Faber and Nierop, 1989). At this point, Gleditsch adopts several coding decisions that we find problematic. First, he assumes that states within a dyad have balanced trade—that the value of goods flowing in one direction is roughly equal to those flowing in another direction (Gleditsch, 2002: 718).[12] Based on this balanced trade assumptions, he replaces missing trade figures with the value of the opposite flow. For Figure 1, this means that we may assume the value of

[11] Among the problems we identified were that Gleditsch (2002): (1) replaced Taiwan's data, all of which were missing from the IMF set, with zero values, suggesting Taiwan had zero trade with the US and other major partners; (2) mistook South Vietnamese trade for North Vietnamese trade; (3) had approximately 3000 observations that were off by a year; (4) approximately 4000 illogical trade value for the United Kingdom; and (5) had incorrect values for his lags and leads, when compared with the actual IMF reports. Gleditsch's Version 4.1 corrects problems 1, 2, and 4.

[12] Our comparisons are based on Gleditsch exptradegdpv4.1, uddtrade_cow.asc.

Table 1. Source Codes and Procedures for COW Trade Data Set 2.0, 1870–2006

Indicator variable code	*Procedure*	*Source*	*Share of observations Flow 1 (% share)*	*Share of observations Flow 2 (% share)*
1	Barbieri Trade Data (1870–1947)	See Barbieri (2002; appendix A)	19,389 (2.99)	19,268 (2.97)
2	Filled in c.i.f. import value	*DOTS* c.i.f. import value (IMF, 2007)	431,254 (66.49)	417,096 (64.30)
3	Filled in missing import value	*DOTS* export value (IMF, 2007)	34,614 (5.34)	49,471 (7.63)
4	Filled in zero import value	*DOTS* export value (IMF, 2007)	14,612 (2.25)	20,801 (3.21)
5	Filled in missing import value	*DOTS* c.i.f. importer value (IMF, 1996)	1,443 (0.22)	1,037 (0.16)
6	Filled in zero import value	*DOTS* c.i.f. importer value (IMF, 1996)	4,403 (0.68)	2,729 (0.42)
7	Filled in missing import value	*DOTS* export value (IMF, 1996)	371 (0.06)	418 (0.06)
8	Filled in zero import value	*DOTS* export value (IMF, 1996)	2,915 (0.45)	616 (0.09)
9	Filled in missing import value	Barbieri value	274 (0.04)	423 (0.07)
10	Filled in zero import value	Barbieri value	1,113 (0.17)	1,320 (0.20)
11	Belgium–Luxembourg data 1948–1996	*DOTS* c.i.f. import value (IMF, 2007)	9,610 (1.48)	3,071 (0.47)
12	Taiwan data	Taiwan mixed sources[1] (1952–1988)	1,150 (0.18)	994 (0.15)
13	Taiwan data	Taiwan Government (1989–2006)	3,311 (0.51)	3,251 (0.50)
–9	Missing	–	124,178 (19.14)	128,142 (19.76)

[1] Taiwan data for 1951–1969 are from United Nations; national data for 1971–1972 are from the APEC Study Center (2008); data for 1973–1988 are from the ROC, Council of Economic Planning and Development (2002, 2004); and data for 1989–2006 are from the ROC, Bureau of Foreign Trade (2008).

US imports from China are identical to China's imports from the US (i.e. Flow 1 is equal to Flow 2, in Figure 1) and use the value of China's imports from the US to replace the US's missing import report. Most scholars realize that assuming China–US trade is balanced is problematic, yet, they may be employing data generated with this decision rule.

Approximately 6% of Gleditsch's data were generated with his balanced trade assumption. This is particularly problematic for conflict scholars who seek to identify and understand the sources of tension in interstate relationships. Trade imbalances are one such source of conflict. By generating data with the problematic balanced trade rule, scholars could erase vital information about why tensions may emerge in a given trading relationship. Ironically, conflict scholars are the only ones we could identify who have used the balanced trade assumption to generate data.

Table 2. The Accuracy of the Balanced Trade Assumption

Balanced trade	*Roughly balanced trade*	*Unbalanced trade*	*No trade*
Flow 1 = Flow 2 1,763	Difference in flows < = 10% 9,129	Difference in flows >10% 178,556	Both flows = 0 92,731

We examined the empirical accuracy of the balanced trade assumption, using dyads with available IMF data for the importer and exporter for two-way trade. To do this, we created a ratio variable that measures the larger to the smaller directional trade flow value. Our findings appear in Table 2. As we see, the majority of trade relationships are unbalanced. Balanced trade is the exception, rather than the rule.

We explore this and other problematic decision rules further by comparing the COW trade data based on IMF reports with the trade data Gleditsch generated, using several decision rules we find to be problematic. Table 3 reveals the dramatic differences between the IMF-based COW Trade figures and Gleditsch's computer generated data. First, Gleditsch generates data for Botswana's imports from the US for the period 1975–1983, using the questionable balanced trade assumption (Gleditsch, 2002: 719). If we look at the period 1975–1983, we see that US imports from Botswana far exceed Botswana's imports from the US and that the balanced trade assumption is inaccurate.

Next, Gleditsch generates additional values for the period 1984–2000, using a second decision rule. He assumes uninterrupted trade—that trade conforms to a particular trend that continues during years with missing data. Gleditsch uses this common, but problematic, assumption to interpolate and extrapolate data for Botswana's imports from the US for the period 1984–2000. Some of these values are generated with data produced with the balanced trade rule discussed above; others use IMF data. Finally, Gleditsch assumes that other years with missing data indicate an absence of trade between the US and Botswana; and he substitutes these values with zeros.

Many scholars assume missing trade data indicates an absence of trade and adopt the decision rule that missing data could be replaced with zeros. Yet, the "missing trade equals zero" trade assumption lacks empirical support and may be problematic for conflict scholars. While some states may engage in no trade, globalization has linked many states that previously had little or no contact. It is also difficult to assume that reports of zero trade are accurate. In the past, the IMF's electronic data files reported zero trade when trade values fell below a minimum threshold. The actual trade values were contained in the IMF paper publications and today appear in both paper and electronic data releases. Yet, today, some states continue to report zero trade in their national accounts when values fall below a certain threshold (UN, 2008).

From a statistical point of view, employing Gleditsch's and other faulty decision rules to generate data to replace missing observations could produce serious measurement error. If scholars use these figures as dependent variables in their analysis

Table 3. Comparing Official and Manufactured Data for US–Botswana Trade

	COW Trade Data		*Gleditsch Trade Data*	
Year	*US–Botswana*	*Botswana–US*	*US–Botswana*	*Botswana–US*
1966	0	–	0	0
1967	0	–	0	0
1968	0	–	0	0
1969	0	–	0	0
1970	0	–	0	0
1971	1	–	0	0
1972	2	2	0	0
1973	2	2	0	0
1974	3	2	3	3
1975	21	2	21	21
1976	54	3	48	48
1977	50.4	2.1	45	45
1978	63.8	2.2	58	58
1979	60.8	5.7	55	55
1980	88.7	6.1	80	80
1981	134.7	6.5	122	122
1982	18.8	5.1	17	17
1983	44.2	4.1	40	40
1984	58.2	18.5	39	40
1985	29.5	16.1	38	40
1986	2.6	20	36	40
1987	7	28.8	35	41
1988	9	41.3	34	41
1989	17.07	30.1	33	41
1990	14.3	19.1	32	41
1991	13.5	30.8	31	41
1992	12.4	46.6	30	42
1993	8.6	24.7	29	42
1994	13.8	22.7	28	42
1995	21.7	35.7	27	42
1996	27.5	28.9	26	42
1997	25.1	–	25	43
1998	20.4	–	20	36
1999	18.3	–	18	35
2000	42	–	42	33

the result will be inefficient estimates; if data are used as independent variables, the estimates will be biased and inconsistent, even as N gets very large (Gujarati, 2003). For these and the other reasons discussed above, we believe it is best to treat missing data as such and to seek alternative sources for trade data and solutions to the missing data problem. We believe some of the faulty decision rules described here may be particularly problematic for scholars who are interested in the periods associated

with intense conflicts and hope to understand how certain independent variables affect and are affected by the onset, duration, and resolution of conflict. If scholars generate data artificially with no consideration of the context of the missing data, particularly in relation to conflict, they are essentially saying that conflict is irrelevant for understanding the trade patterns or that trade patterns exist independent of conflict. Ironically, some scholars use such trade data to investigate the relationship between trade and conflict, when their assumptions often suggest no relationship exists.

As mentioned, we believe trade data are more likely to be missing during periods of intense conflict or in highly conflictual relationships. Given the importance of having accurate trade data for periods associated with intense conflict, one should strive to develop accurate assumptions about missing data and should seek multiple sources for data, rather than limiting their search. It is important to consider how accurate the trade data are for analyzing periods of conflict and whether data are more likely to be missing during these periods or in conflictual relationships. If such a pattern exists and data are being manufactured with little or no attention to conflict, the consequences could be quite serious.

Trade and Conflict

One important question is whether the problems we outline and alternative strategies to address them may produce to different empirical findings. It is difficult to answer this question definitively. First, trade data are used for a variety of purposes and the impact of artificially produced data might vary, depending upon one's total sample size and the portion of the sample that relies upon manufactured data. For example, someone might want to examine one dyad over time and might not realize that the majority of their time series data were produced with problematic assumptions and not obtained from actual trade reports. Imagine a case in which states were experiencing an upward trend in trade over time, but entered a period of intense conflict in which they failed to report trade. Suppose further that this trade returns to the pre-conflict level once reporting begins again. If we replace the missing trade data using a linear trend, without considering the context in which data are missing, we are likely to produce inaccurate reports. Unfortunately, we often employ data without fully understanding how it was generated or its limitations. While we may not know how alternative data strategies impact our findings, we are able to assess how frequently data are missing during periods of conflict and whether some scholars tend to rely upon questionable data for these critical observations. We also examine whether dyads with missing data are different in their conflict propensities than dyads with available data. This test also reveals whether it is reasonable to use standard assumptions to replace missing data points. To explore the relationship between conflict and trade data availability, we perform some preliminary analyses.

First, we examine the extent to which data are missing during conflict periods. Next, we examine whether dyads with missing data are more or less likely to engage in conflict. In the first analysis, we compare the COW Trade Data Set with Gleditsch's trade data and examine the patterns of missing and manufactured data. We use the COW Militarized Interstate Dispute (MID) data to divide our sample into years in which a dyad experienced a conflict and those in which it did

Table 4. Trade Data Availability and MIDs

	COW Trade		*Gleditsch Trade*	
	Missing	*Not missing*	*Artificial data*	*Not missing*
MIDs	145	2,176	463	1,798
	6.27%	93.73%	20.48%	79.52%
No MIDs	61,601	487,457	231,735	293,306
	11.22%	88.78%	44.14%	55.86%

not (Jones et al., 1996; Ghosn, et al., 2004; and Ghosn and Bennett, 2003). [13] We then examine the frequency with which trade data are missing for conflict years; the availability of official trade reports; and the extent to which some scholars rely upon trade data generated with the problematic assumptions we critiqued above. Our results appear in Table 4.

We see that for the COW Trade Data Set, trade data are missing for approximately 6% of the MID observations and about 11% of non-MID observations. For the Gleditsch data set, more than 20% of the MID observations and 44% of non-MID observations consist of computer generated trade values. As mentioned, if data are missing as a result of conflict, it might be particularly problematic to apply some of the assumptions we described above. The fact that Gleditsch has made assumptions about the trade values for approximately 20% of all MIDS could affect any conflict analysis that relies upon this data set. The question of how much the results are affected is less clear, since official statistics are not always available for these observations. Yet, it is still possible to evaluate whether dyads with missing data are somehow different than those without missing data.

Next, we examine whether dyads with missing data have a different propensity toward conflict than those without missing data.[14] If they do, the way we handle missing data may be even more important than some believe. To investigate this issue, we perform a set of statistical analyses that examines the probability of conflict in dyads with missing data and compare this to cases with available data. We also compare dyads with reports of zero trade to those with positive trade values, since replacing missing trade data with zero values is a common way of addressing the missing data problem.

Our goal here is not to resolve the question of which model is best for analyzing the trade–conflict relationship. Instead, we simply wish to make comparisons about the conflict propensity of different dyad types, based upon trade data availability, rather than trade level. Thus, our measure of trade in a basic trade–conflict analysis related to trade data type and not level.

Our model includes the few variables present in most trade–conflict analyses—conflict, trade, GDP, and geographic proximity—and includes all non-directed dyads

[13] We utilize EUGene Version 3.023 to generate the MID data and the DISTANCE data described below (Bennett and Stam, 2000, 2007).

[14] Future analysis will consider the simultaneous nature of the trade–conflict relationship.

Table 5. Conflict Analysis for Dyads with Missing, Zero, and Available Data

		MIDs		*Fatal MIDs*		*Wars*	
		Onset	*Involvement*	*Onset*	*Involvement*	*Onset*	*Involvement*
IMF	Missing	0.059***	0.121***	–0.286***	–0.283***	0.271***	0.409***
	Zero	–0.074**	–0.126***	–0.104***	–0.105***	0.199**	–0.109*
COW	Missing	0.054*	0.072***	–0.022***	–0.021***	0.336***	0.228***
	Zero	–0.023	0.027	–0.211***	–0.210***	0.194***	0.206***
Gleditsch	Missing	–0.141***	–0.101***	–2.765***	–2.597***	.0297***	0.293***
	Zero	–0.127***	0.157***	–3.261***	2.987***	–0.051***	–0.232***

*$p < .10$; **$p < .05$; ***$p < .01$.

for the years 1948–2001. Our dependent variable is conflict and we measure that using three different types of MIDs: all MIDs, FATAL MIDs (MIDs with at least 1 fatality), and WARS (MIDs with more than 1,000 battle deaths). We also examine MID ONSET and MID INVOLVEMENT. MID ONSET is coded 1 for the first year of the dispute and 0 otherwise. MID INVOLVEMENT is coded 1 for each year of the conflict and coded 0 during years of peace.

We use three different data sets in our analysis: (1) the original IMF dyadic data without the COW adjustments (IMF); (2) the COW Trade Data Set (COW); and (3) the Gleditsch Trade Data Set (Gleditsch, 2002). For the central variable of interest, trade data, we create two dichotomous variables per data set. The first is MISSING DATA and is set to 1 if the trade values are missing and 0 otherwise. For Gleditsch, we set his computer generated values to 1, since we assume these values were originally missing.[15] For the second variable, ZERO TRADE, the dummy variable is coded as 1 if dyadic trade equals zero and is coded 0 for non-zero trade values. For Gleditsch, the dummy variable includes IMF zeros and his assumed zeros.

Our model includes the most common control variables found in trade–conflict analysis: distance and GDP. DISTANCE is measured as capital to capital distance with adjustments for contiguity.[16] GDP data come primarily from the World Bank (2008) and were converted to real values using a conversion factor index (Sahr, 2008). DISTANCE is assumed to be negatively related to conflict, while GDP is assumed to be positively associated with conflict. These variables were both statistically significant and had the expected signs. For ease of presentation, we do not include the control variables in Table 5, where we present the results of our analysis.

We see that for the IMF and COW Trade Data Sets, dyads with MISSING DATA tend to be more likely to engage in MIDs and WARs and less likely to experience FATAL MIDs. For Gleditsch, dyads with MISSING DATA are less likely to experience

[15] For Gleditsch, this holds when variables giabo & gibao are greater than 2.

[16] Gleditsch combines trade and GDP into one ratio variable, which was not appropriate for this analysis.

MIDs and FATAL MIDs, but more likely to experience WARS. The findings are statistically significant, which means that there are significant differences in the conflict propensity of dyads with missing trade data and those with available data. The differences are not consistent across conflict types, but we believe the results provide sufficient evidence to question the usual assumptions about trade values when data are missing, particularly if we hope to use these data to analyze conflict.

Our results comparing the conflict propensity of dyads with ZERO TRADE and those with positive trade also vary across data sets. For the IMF, dyads with ZERO TRADE are less likely to engage in conflict. The only exception is WAR ONSET, where dyads with no trade are more likely to witness an outbreak of war. For COW Trade, dyads with ZERO TRADE are less likely to engage in MIDs, but more likely to see the outbreak of war. Finally, Gleditsch's ZERO TRADE dyads are negatively associated with MID and FATAL MID ONSET and positively associated with MID and FATAL MID INVOVLEMENT. However, ZERO TRADE is negatively associated with the predicted probability of both WAR ONSET and INVOLVEMENT. Given the variation in results for cases with ZERO TRADE, it seems unwise to make assumptions about when ZERO TRADE values should be used in place of missing data and when reports of no trade are accurate.

Overall, the results substantiate the concerns raised in this paper: that replacing missing trade data with questionable values is problematic and may affect our results, particularly when we do not explore the reasons the data are missing. Missing trade data tend to be associated with significant increases or decreases in different types of conflict. We would want to know if missing data coincided with conflict in a relationship. In addition, conflict might be short or long-term; it may or may not affect trade and the reporting of trade; and trade may or may not vary over time. Furthermore, we have only discussed direct trade and conflict and have not considered how conflict might affect trade relations with third parties (e.g. allies and adversaries of each state in the conflict).

If we conclude that missing trade data indicates certain values of trade, then we assume away other conditions. The assumptions we make could determine the results we obtain and produce tautologies. For example, if we assume dyadic trade conforms to a particular trend and use that to predict conflict, we might produce values that determine our findings about trade and conflict. The assumption that there is a trend in the data will impact the results and, therefore, the results are driven by the assumption about missing trade values. This is clearly not how we want to study the trade–conflict relationship or many other facets of trade.

Furthermore, our results should not be used to justify filling in missing trade values, because there are situations where missing trade data may be positively or negatively related to conflict. Findings of a negative result between conflict and trade may be overestimated, if we assume missing data are zero and the missing cases tend to be more conflict prone. Findings of a positive relationship between missing trade and conflict may be underestimated, since we do not know whether trade was going up or down or was nonexistent before, during, and after the conflict. Thus, filling in missing trade data using assumptions is not a solution to the underestimation or overestimation problem. In fact, a finding that filling in missing values increases the strength of the pre-filled-in missing results, as some contend (Gleditsch, 2002), is not a finding at all, since we have

no information about why the trade values are missing. Filling in missing trade values becomes even more problematic if conflict is impacting trade. For all these reasons, we are dubious about current methods for filling in missing trade values.

Conclusions

This paper reveals the importance of understanding the trade data that we employ in our research. The issue of missing data and how it impacts our analysis is one that must be addressed with greater care than conflict researchers have done thus far. While some have sought remedies to the problem of missing data, we believe no one has found the optimal solution. Until that solution comes, we must consider how missing data patterns affect the answers we receive about questions that involve international trade.

More importantly, when data are generated with problematic decision rules, it is difficult to have confidence in the results they generate. Our goal should be to devise better strategies to obtain valid and reliable data. To this end, we believe that the new Correlates of War Trade Data Set is a first step in this process. Future efforts need to move beyond the realm of IMF trade statistics and official national reports. We must seek sources and, when possible, build data-based measures that capture the most accurate picture of world trade, including the actors involved, the goods they supply and demand, the linkages and dependencies that exist across countries, the means by which laws are circumvented, and the ways trade data are falsified or distorted.

We would prefer to see more cooperation among scholars interested in trading relationships. When dealing with extraordinarily large numbers of observations, the possibility of machine-induced coding errors is greater than with smaller data sets where cross-checking of every individual observation is practical. We hope that researchers employing the new Correlates of War Trade Data Set will report any problems they discover to its creators. Any trade data set will benefit from input from a larger community of scholars. We have identified and begun working on several issue areas that we believe will augment our understanding of global trade relations and their myriad effects upon the lives of nations.

As mentioned, our goal is to provide an accurate depiction of trading relationships. Yet, as is true of any quantitative indicators, we must recognize that our measures may paint a picture that stands in sharp contrast from reality. We must understand the shortcomings of certain research strategies, as we seek solutions to our data problems and interpret empirical findings tied to different data sets. Future research should be better able to produce valid, reproducible findings if we generate more reliable data on interstate trade. While this article and the trade data project it describes strive to bring us one step closer to meeting our scientific objectives of properly measuring trading relationships, it is by no means the final step.

References

APEC Study Center, City University of Hong Kong Data Bank. 2008. http://fbweb.cityu.edu.hk/hkapec/DataBank/Taiwan-Trade-Balance.pdf.

Barbieri, Katherine. 2002. *The liberal illusion: Does trade promote peace?* Ann Arbor: University of Michigan Press.

Barbieri, Katherine, and Swapna Pathak. 2007. Free trade and terrorism. In *Terrornomics,* eds Sean S. Costigan and David Gold, ch. 7. Burlington, VT: Ashgate.

Barbieri, Katherine, Omar Keshk, and Brian M. Pollins. 2003. The Correlates of War Trade Data Set: A progress report. Paper presented at the Annual Meeting of the Peace Science Society (International), Ann Arbor, Michigan, 13–16 November.

Bennett, D. Scott, and Allan Stam. 2000. *EUGene*: A conceptual manual. *International Interactions* 26:179–204. Website: http://eugenesoftware.org.

Bennett, D. Scott, and Allan Stam. 2007. EUGene: Expected Utility Generation and Data management program. Version 3.203, 29 October. http://www.eugenesoftware.org/.

Bhagwati, Jagdish N. 1964. "On the Underinvoicing of Imports." *Oxford Bulletin of Economics and Statistics* 27(4): 389–397.

Bhagwati, Jagdish N. 1967. Fiscal policies, the faking of foreign trade declarations, and the balance of payments. *Oxford Bulletin of Economics and Statistics* 29(1): 61–77.

Bhagwati, Jagdish N., and T. N. Srinivasan. 1981 The evaluation of projects at world prices under trade distortions: Quantitative restrictions, monopoly power in trade and nontraded goods. *International Economic Review* 22(2): 385–399.

Correlates of War Project. 2008. State System Membership List, v2008.1. Online, http://correlatesofwar.org.

de Wulf, Luc. 1981. Customs valuation and the faking of invoices. *Economia Internazionale* 34(1): 13–33.

Ely, J. Edward. 1961. Variations between U.S. and its trading partner import and export statistics. *American Statistician* 15(2): 23–26.

Faber, Jan, and Tom Nierop. 1989. *World export data, 1948–1983* (ICPSR no. 9116). Amsterdam: University of Amsterdam (producers); Ann Arbor, MI: Inter-university Consortium for Political and Social Research (distributors).

Ghosn, Faten, and Scott Bennett. 2003. Codebook for the Dyadic Militarized Interstate Incident Data, Version 3.10. Online: http://cow2.la.psu.edu

Ghosn, Faten, Glenn Palmer, and Stuart Bremer. 2004. The MID3 Data Set, 1993–2001: Procedures, coding rules, and description. *Conflict Management and Peace Science* 21: 133–154.

Ghosn, Faten, and Scott Bennett. 2003. Codebook for the Dyadic Militarized Interstate Incident Data, Version 3.10. Online: http://cow2.la.psu.edu.

Gleditsch, Kristian S. 2002. Expanded trade and GDP data. *Journal of Conflict Resolution* 46: 712–24.

Gujarati, Damador N. 2003. *Basic econometrics*. Boston, MA: McGraw Hill.

International Monetary Fund. 1993. A guide to direction of trade statistics (http://www.imf.org/external/pubs/cat/longres.cfm?sk=154.0).

International Monetary Fund. 2006. *Direction of trade statistics: Historical database and browser* on CD-ROM.

International Monetary Fund. 2007. *Direction of trade statistics on CD-ROM*. Monthly Subscription.

International Monetary Fund. 2008. Data and statistics: Direction of trade statistics consistency between partner country data (http://www.imfstatistics.org/dot/DOTConsi.htm).

Jones, Daniel M., Stuart A. Bremer, and J. David Singer. 1996. Militarized interstate disputes, 1816–1992: Rationale, coding rules, and empirical patterns. *Conflict Management and Peace Science* 15(2): 163–213.

Keshk, Omar M. G., Brian M. 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: 1155–1179.

Mansfield, Edward D., and Brian M. Pollins, eds. 2003. *Economic interdependence and international conflict: New perspectives on an enduring debate*. Ann Arbor: University of Michigan Press.

Morgenstern, Oskar. 1963. *On the accuracy of economic observations*. Princeton, NJ: Princeton University Press.

Morgenthau, Hans J. 1948. *Politics among nations: The struggle for power and peace*. New York: Knopf.

Oneal, John R., and Bruce M. Russett. 1997. The classical liberals were right: Democracy, interdependence, and conflict, 1950–85. *International Studies Quarterly* 41(2): 267–294.

Pak, Simon J., and John S. Zdanowicz. 1994a. An estimate of lost US federal income tax revenues due to over-invoiced imports and under-invoiced exports. Report 9404–01, Trade Research Institute, Miami, FL.

Pak, Simon J., and John S. Zdanowicz. 1994b. A statistical analysis of the US merchandise trade data base and its uses in transfer pricing compliance and enforcement. *Tax Management Transfer Pricing Report* 3(1) 50–57.

Pak, Simon J., Stelios H. Zanakis, and John S. Zdanowicz. 2003. Detecting abnormal pricing in international trade: The Greece–USA Case. *Interfaces* 33(2): 54–64. Linthicum, MD: Institute for Operations Research and the Management Sciences (INFORMS).

Republic of China (ROC), Bureau of Foreign Trade Website. 2008. Data source, Directorate General of Customs, Ministry of Finance (http://cus93.trade.gov.tw/english/FSCE/FSC0011E.ASP).

Republic of China (ROC), Council for Economic Planning and Development. 2002, 2004. *Taiwan Statistical Data Book* (http://www.cepd.gov.tw/english/file/2004DataBook.pdf?CFID=87002&CFTOKEN=50209707).

Russett, Bruce, and John Oneal. 2001. *Triangular peace: Democracy, interdependence, and international organizations*. New York: W. W. Norton.

Sahr, Robert. 2008. Inflation conversion factors for dollars 1774 to estimated 2018. Corvallis, OR: Oregon State University. (http://oregonstate.edu/cla/polisci/faculty-research/sahr/sahr.htm).

Schneider, Gerald, Katherine Barbieri, and Nils Petter Gleditsch. 2003. Does globalization contribute to peace? A critical survey of the theoretical and formal literature. In *Globalization and armed conflict*, eds Gerald Schneider, Katherine Barbieri and Nils Petter Gleditsch, pp. 3–29. Lanham, MD: Rowman & Littlefield.

Sheikh, Munir. 1974. Underinvoicing of imports in Pakistan. *Oxford Bulletin of Economics and Statistics* 36(4): 287–296.

United Nations, Statistical Office, Department of Economic Affairs. (1951–1984). *Yearbook of international trade statistics.* New York: United Nations.

United Nations. 2008. International merchandise trade statistics national compilation and reporting practices. (UN Website http://unstats.un.org/unsd/tradereport/question.asp).

World Bank. 2005, 2008. *World development indicators*. On-Line Version. Washington, DC: World Bank.

Yeats, Alexander. 1978. On the accuracy of partner country trade statistics. *Oxford Bulletin of Economics and Statistics* 40(1): 341–361.

Yeats, Alexander. 1990. On the accuracy of economic observations: Do Sub-Saharan trade statistics mean anything? *World Bank Economic Review* 4(2): 135–156.

KATHERINE BARBIERI is Associate Professor of Political Science at the University of South Carolina and author of *The Liberal Illusion: Does Trade Promote Peace?* Her data set on trade and conflict is available to scholars and she now serves as a co-host for the Correlates of War Trade Data Set.

OMAR M. G. KESHK is senior lecturer in the Political Science Department and the Undergraduate International Studies Program at the Ohio State University. He has published in the *Journal of Politics*, *Conflict Management and Peace Studies*, and the *Stata Journal*. He co-hosts the COW Trade Data Set.

BRIAN M. POLLINS is Emeritus Associate Professor of Political Science at the Ohio State University. He now teaches and writes at the Naval Postgraduate School in Monterey, CA.